

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ANALYSE DES PRATIQUES EN ÉVALUATION DE PROGRAMME : LES ÉLÉMENTS
QUI CONTRIBUENT À FORMULER UN JUGEMENT FONDÉ

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉDUCATION

PAR
STÉPHANIE MONGIAT

OCTOBRE 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Avant de présenter mon projet de mémoire, je souhaite remercier chaleureusement tous ceux l'ayant rendu possible.

Merci tout d'abord au professeur Marthe Hurteau (directrice de ce mémoire et directrice du projet de recherche à l'intérieur duquel celui-ci s'inscrit). Elle s'est avérée beaucoup plus qu'une directrice, voire un mentor à mon égard, tout en m'introduisant au domaine de la recherche et de l'évaluation de programme. Je me considère privilégiée d'avoir été accompagnée par un professeur si compétent et professionnel.

Merci au docteur Suzanne Véronneau-Troutman qui m'a accordé une bourse à la mémoire de Denise Véronneau, professeur en éducation à l'UQÀM. Merci au professeur Johanne Otis (Titulaire de la Chaire de recherche du Canada en éducation à la santé) dont les précieux conseils et l'appui financier (attribution d'une bourse d'étude) furent aussi très appréciés. Merci au professeur Juan Wood dont l'accompagnement fut sans pareil durant l'intégralité de mon cheminement. Merci au professeur Sylvain Houle (professeur membre de l'équipe de recherche) qui a fait preuve d'une grande générosité en partageant ses réflexions quant à la nature et la portée du projet de recherche. L'enthousiasme et le dévouement de ces professeurs ont indéniablement contribué aux mérites de ce projet.

Merci également à mes collègues, Isabelle Grenier et Lina Zielinski, dont la contribution tant morale qu'intellectuelle fut très appréciée. Elles furent de véritables complices au cours des différentes étapes d'une maîtrise. C'est certainement avec fierté que nous devons regarder le chemin parcouru ensemble.

Enfin, mes dernières pensées vont tout naturellement à ma famille. Elles vont à ma mère, Maureen Fortin, ma fidèle et fervente correctrice. Elles vont à mon père, Vital Mongiat, le premier à m'avoir encouragé à entamer ce projet. Elles vont finalement à mon conjoint, Marc-André Denoncourt, qui s'est assuré que j'avais tout le soutien nécessaire pour persévérer dans les moments d'incertitude.

TABLE DES MATIÈRES

LISTE DES FIGURES	vi
LISTE DES TABLEAUX	vii
RÉSUMÉ.....	viii
INTRODUCTION.....	1
CHAPITRE I	
PROBLÉMATIQUE	3
1.1 L'évaluation de programme	3
1.2 La controverse qui existe au sein de l'évaluation de programme.....	5
1.3 Problématique de la recherche	7
1.4 La question de recherche.....	8
1.5 La pertinence de l'étude	9
1.5.1 La pertinence scientifique.....	9
1.5.2 La pertinence sociale	10
1.6. Synthèse	10
CHAPITRE II	
CADRE THÉORIQUE.....	12
2.1 Distinction entre les théories descriptives et normatives.....	12
2.2 Le processus spécifique à l'évaluation de programme	14
2.2.1 La contribution de Scriven (1980): la logique de l'évaluation	15
2.2.2 La contribution de Hurteau (1991): les choix stratégiques et méthodologiques.....	16
2.2.3 La contribution de Fournier (1995): la logique de travail.....	17
2.2.4 La contribution de Stake (2004): la pensée interprétative	18
2.2.5 La contribution d'Arens (2006) : la pondération des critères	19
2.2.6 La contribution de Hurteau et Houle (2006) : une modélisation	20
2.3 Les constituants de la modélisation.....	22
2.3.1 La description du programme	22
2.3.2 La problématique de l'évaluation	24
2.3.3 L'opérationnalisation de l'évaluation	25

2.3.4 Les choix stratégiques de l'évaluation.....	26
2.4 Le jugement fondé.....	28
2.5 Synthèse et objectifs de recherche.....	31
CHAPITRE III	
LA DÉMARCHE MÉTHODOLOGIQUE.....	32
3.1 Le type de recherche.....	32
3.2 Les choix méthodologiques.....	33
3.2.1 Le type d'information requise.....	33
3.2.2 Le choix de l'échantillon	33
3.2.3 La collecte des données	34
3.3 L'analyse des données.....	37
3.4 Les forces de l'étude.....	38
3.5 Les limites de l'étude	40
3.6 Synthèse	41
CHAPITRE IV	
LES RÉSULTATS	42
4.1 Objectif 1 : L'application de la modélisation	45
4.1.1 La description du programme	45
4.1.2 La problématique de l'évaluation	46
4.1.3 L'opérationnalisation de l'évaluation	47
4.1.4 Les choix stratégiques.....	49
4.1.5 Conclusion (Objectif 1)	52
4.2 Objectif 2 : Les caractéristiques des jugements et de leurs fondements.....	53
4.2.1 Le jugement et ses fondements; d'après les constats de Arens (2006).....	54
4.2.2 Le jugement et ses fondements; d'après le modèle de Toulmin (1993)	56
4.2.3 Conclusion (Objectif 2)	59
4.3 Objectif 3 : Étudier la contribution de l'application de la modélisation à la formulation d'un jugement fondé.....	61
4.3.1 La mise en relation d'après les analyses quantitatives.....	62
4.3.2 L'analyse de la contribution selon un examen qualitatif	62
4.3.3 Conclusion (Objectif 3)	64

CHAPITRE V	
LA DISCUSSION	65
5.1 Quels sont les résultats les plus révélateurs de la pratique actuelle en évaluation de programme (objectif 1)?	65
5.2 Dans quelle mesure les résultats concernant les caractéristiques des jugements et de leurs fondements (objectif 2) précisent-ils le malaise énoncé par Guba (1972) et Scriven (1995)?	68
5.3 Quel est l'impact des résultats concernant la contribution de la modélisation à la formulation d'un jugement fondé (objectif 3) dans le domaine de l'évaluation de programme?	70
CONCLUSION	72
APPENDICE A	
LE DÉTAIL DES RAPPORTS D'ÉVALUATION RETENUS	74
APPENDICE B	
GRILLES D'ANALYSE	77
RÉFÉRENCES	83

LISTE DES FIGURES

Figure	Page
2.1	Le jugement: une synthèse pondérée de différentes déclarations selon Arens (2006)..... 19
2.2	La modélisation du processus spécifique à l'évaluation de programme et les fondements théoriques de ses composantes.(Hurteau et Houle, 2006)..... 21
2.3	Un programme: un système. (Tirée de Chen, 2005)..... 23
2.4	Les indicateurs témoignant d'un jugement fondé inspirés des constats d'Arens (2006) incorporés à sa représentation de la construction d'un jugement..... 29
2.5	Les indicateurs témoignant d'un jugement fondé inspirés de la contribution de Toulmin (1993) incorporés à la représentation de la construction d'un jugement d'Arens (2006)..... 30
4.1	La modélisation du processus spécifique à l'évaluation de programme: éléments et interrelations. (Hurteau et Houle, 2006)..... 43
4.2	Les quatre premières composantes de la modélisation sont développées au premier objectif..... 44
4.3	Les jugements, la finalité de la modélisation, sont caractérisés au deuxième objectif. 53
4.4	Les indicateurs témoignant d'un jugement fondé inspirés des constats d'Arens (2006) incorporés à sa représentation de la construction d'un jugement..... 54
4.5	Les indicateurs témoignant d'un jugement fondé inspirés de la contribution de Toulmin (1993) incorporés à la représentation de la construction d'un jugement d'Arens (2006)..... 56
4.6	La distribution des cas selon les caractéristiques des jugements..... 60
4.7	La relation à étudier au troisième objectif. 61

LISTE DES TABLEAUX

Tableaux	Page
2.1 Distinction entre les théories descriptives et normatives	13
2.2 Relation entre les questions (et/ou objectifs) et le type d'évaluation selon Rossi, Lipsey et Freeman (2004)	26
3.1 Les indicateurs des grilles d'analyse témoignant de l'application de la modélisation	35
3.2 Les indicateurs de la grille d'analyse caractérisant le jugement et ses fondements	36
4.1 Analyse du contenu lié à la <i>Description du programme</i>	45
4.2 Analyse du contenu lié à la <i>Problématique de l'évaluation</i>	46
4.3 Analyse du contenu lié à l' <i>Opérationnalisation de l'évaluation</i>	48
4.4 Analyse du contenu lié aux <i>Choix stratégiques de l'évaluation</i>	50
4.5 Fréquence moyenne des indicateurs des composantes liées au premier objectif de recherche	52
4.6 Analyse du contenu lié au <i>jugement et ses fondements</i> d'après l'apport d'Arens (2006)	55
4.7 Analyse du contenu lié au <i>jugement et ses fondements</i> d'après l'apport de Toulmin (1993)	57
4.8 Examen des résultats qualitatifs des jugements pour y déceler si des éléments de la modélisation y sont repris	63

RÉSUMÉ

Grâce à son caractère indéniablement pratique ainsi qu'au développement de son corpus théorique, l'évaluation de programme s'est imposée à l'intérieur des diverses organisations chargées de la gestion des programmes éducatifs et sociaux. Paradoxalement, ce domaine est également périodiquement remis en question par les utilisateurs. Ils doutent de sa capacité à remplir ses fonctions premières, soit produire des jugements suffisamment crédibles pour être acceptés des clients et ainsi, guider les grandes orientations des programmes. Pour expliquer ce phénomène, Guba (1972) et Scriven (1995) avancent que ce serait une mauvaise application du processus spécifique à l'évaluation de programme qui nuirait à la formulation d'un jugement fondé. Leurs propos demeurant cependant de l'ordre des intuitions, la présente étude vise à approfondir la contribution de l'application du processus spécifique à l'évaluation de programme à la formulation d'un jugement fondé. Les pratiques telles que rapportées ou écrites à l'intérieur des rapports d'évaluation (40) sont ainsi analysées afin de cerner cette contribution. Plus précisément, cette étude poursuit les efforts investis au premier volet du projet de recherche (Hurteau et Houle, 2006) sur les pratiques évaluatives et appuie son corpus théorique sur l'apport de Scriven (1980), de Hurteau (1991), de Fournier (1995), de Stake (2004) et d'Arens (2006).

Les résultats montrent que *l'Opérationnalisation de l'évaluation* qui spécifie le type d'évaluation, les objectifs et/ou les questions d'évaluation n'est pas réalisée de manière systématique et présente des irrégularités au niveau de son utilisation. Ils révèlent également qu'il n'y a que la moitié des rapports étudiés qui ont su dépasser les évidences pour produire un jugement et qu'il y en a encore moins qui les nuancent par les limites ou forces du devis de l'évaluation. Ce constat met en doute la teneur même des rapports d'évaluation actuellement produits. Ces rapports constituent-ils réellement des évaluations ou sont-ils davantage de l'ordre d'une simple enquête? Ce constat s'ajoute au fait qu'aucun rapport n'explique la manière dont les déclarations ont été synthétisées pour construire le jugement et que plusieurs rapports émettent des recommandations qui dépassent la portée du type d'évaluation effectué. Enfin, les résultats suggèrent qu'il n'y aurait pas de relation significative entre l'application de la modélisation et la formulation d'un jugement fondé. Un examen plus qualitatif met toutefois en évidence que certains éléments de la modélisation, hormis ceux liés à *l'Opérationnalisation de l'évaluation*, peuvent être réinvestis à l'intérieur d'une argumentation pour convaincre ou dissuader du bien-fondé du jugement. Enfin, cette recherche cerne avec plus de précision le malaise touchant la connaissance et/ou application du processus spécifique à l'évaluation de programme et la formulation d'un jugement fondé dans le domaine de l'évaluation de programme.

Mots clefs :

Évaluation de programme, logique de l'évaluation, pratiques évaluatives, jugement.

INTRODUCTION

Au cours des dernières décennies, l'évaluation de programme s'est imposée à l'intérieur des diverses organisations chargées de la gestion des programmes éducatifs et sociaux grâce à son caractère pratique ainsi qu'au développement de son corpus théorique. L'arrivée de revues spécialisées et la tenue de congrès internationaux mettent également en évidence que ce domaine est en pleine expansion autant au Canada qu'ailleurs dans le monde.

Paradoxalement, plusieurs auteurs dont King (2003) dans son article intitulé *The Challenge of Studying Evaluation Theory*, sont d'avis que l'évaluation de programme ne possède pas, à ce jour, assez de «maturité» pour offrir aux évaluateurs un ancrage théorique lui permettant d'encadrer adéquatement sa pratique. À cet effet, ceux-ci s'inquiètent de son avenir et soulignent que la majorité de ses fondements théoriques se basent sur des spéculations concernant ce qui devrait être fait plutôt que sur des données scientifiques. Ces constats seraient également observés en pratique, où la capacité de l'évaluation de programme à remplir ses fonctions premières serait également remise en cause.

Actuellement, ce domaine déploie des efforts indéniables dans l'élaboration de son propre corpus théorique pour remplir efficacement sa fonction première, à savoir produire un jugement suffisamment crédible pour être accepté des clients et guider les grandes orientations des programmes. La présente recherche s'inscrit à l'intérieur d'un programme de recherche qui vise à cerner les éléments qui définissent et fondent l'acte d'évaluer à l'intérieur de ce domaine. Le premier volet de la recherche ayant permis l'élaboration d'une modélisation du processus spécifique à l'évaluation de programme, le second, à l'intérieur duquel cette recherche s'inscrit, se fixe comme principal objectif d'approfondir sa contribution à la formulation d'un jugement fondé, raison d'être de toute démarche évaluative. L'ensemble de cette démarche vise à contribuer au développement global du domaine de l'évaluation de programme.

Ainsi, le présent document présente l'ensemble de la démarche scientifique effectuée au volet II du programme de recherche. Plus précisément, il approfondit la problématique à l'intérieur du chapitre premier pour ensuite tracer, au deuxième chapitre, un portrait théorique des développements la concernant. C'est au troisième chapitre que sont abordées les différentes précisions méthodologiques liées à la réalisation de cette recherche suivies d'une présentation des résultats au quatrième chapitre. Enfin, les résultats font l'objet d'une discussion au cours du dernier et cinquième chapitre.

CHAPITRE I

PROBLÉMATIQUE

Ce premier chapitre est consacré à établir la problématique sur laquelle se penche cette recherche. Pour ce faire, l'état actuel du domaine de l'évaluation de programme y est décrit et le malaise auquel ce domaine est confronté y est présenté. Enfin, il se termine en présentant la question de recherche ainsi que la pertinence, tant sociale que scientifique, de cette étude.

1.1 L'évaluation de programme

Le développement théorique de l'évaluation de programme étant relativement récent, ses enjeux sont, encore aujourd'hui, étroitement liés à ceux des différents domaines qui ont contribué à son évolution. Selon cette perspective, il apparaît pertinent de tracer un bref historique de son développement pour le conclure par sa définition actuelle.

L'être humain a toujours eu besoin de mesurer et c'est probablement ce qui a motivé, dès l'Antiquité, les empereurs chinois à entreprendre les premières évaluations (Guba et Lincoln, 1989). C'est d'ailleurs grâce à ce besoin d'une mesure toujours plus utile et précise que se sont développées les sciences de la mesure, des statistiques et des tests. Il est important de le mentionner, car elles ont contribué à la première définition du terme « évaluer », c'est-à-dire associer une valeur à un objet, appelé « objet » d'évaluation en évaluation.

Ce n'est cependant qu'à partir de 1850 qu'est arrivé le besoin systématique d'évaluer les effets d'une intervention. En effet, c'est à cette période que les politiques gouvernementales américaines concernant les conditions sociales de la population sont mesurées pour la première fois (Weiss, 1998). En mesurant la différence entre les objectifs et leurs résultats en pratique, les statisticiens ont été les premiers à effectuer des évaluations de programme qui ont réellement influencé les décisions politiques.

C'est toutefois le contexte des années 60 qui permet à l'évaluation de programme de prendre un réel essor. Alors que plusieurs programmes sont mis en place par l'état américain pour contrer la pauvreté suivant la Deuxième Guerre mondiale, le gouvernement investit en masse pour l'évaluation de l'efficacité et du rendement des organisations dépendantes des fonds publics (Madaus et Stufflebeam, 2004). L'évaluation de programme est alors si près du domaine de la recherche en sciences sociales que ces deux domaines peuvent facilement être confondus.

À partir des années 70 apparaissent une variété de revues scientifiques spécialisées en évaluation de programme ainsi que de nombreuses publications qui enrichissent son corpus théorique. Elles favorisent la professionnalisation de ce domaine (Rossi, Lipsey et Freeman, 2004). L'évaluation de programme se distingue alors tranquillement de la recherche en sciences sociales par son caractère plus près des pratiques concernant la gestion des programmes sociaux. Ainsi, contrairement à la recherche qui vise à comprendre les lois ou les principes qui régissent notre monde, elle se distingue en visant plutôt à produire un jugement concernant un aspect d'un programme (implantation, résultats,...) ou encore ses interrelations (efficience, efficacité, ...), permettant aux clients de mieux le saisir et d'éclairer les prises de décision concernant ses grandes orientations.

Bien qu'elle se serve toujours des sciences de la mesure et de la théorie derrière les programmes sociaux, l'évaluation de programme est désormais intimement liée à un acte spécifique, à savoir celui de *juger*, ce qui fait qu'elle implique nécessairement des valeurs. C'est ainsi que, pour orienter les valeurs à préconiser à l'intérieur de la pratique de

l'évaluation de programme, plusieurs auteurs ont créé des modèles d'évaluation constitués principalement de leurs recommandations à cet effet.

Alors que l'évaluation de programme ne consistait, au point de départ, qu'à « mesurer » les effets d'un programme en terme statistique, ce domaine s'est graduellement transformé en un domaine distinct de la recherche par sa spécificité (Scriven 2006). Tous ces développements enrichiront ce que signifie l'évaluation de programme qui pourrait enfin se définir aujourd'hui comme une cueillette systématique de données ayant pour objet d'étude les différents aspects d'un programme ainsi que leurs interrelations, et ce, en vue de porter un jugement de valeur suffisamment crédible aux yeux des clients de l'évaluation pour guider les grandes orientations de ces programmes. Cette définition s'inspire de nombreux auteurs tels que House (1980), Fournier (1993), Weiss (1998), Stake (2004), Chen (2005), Davidson (2005) et Scriven (2006).

En ce qui concerne l'évaluation de programme au sein du domaine de l'éducation, son rôle est primordial. En effet, les plus grands programmes éducatifs tels que les programmes de formation de l'école québécoise du niveau primaire et secondaire, sont évalués lors de leurs différentes phases d'implantation pour guider leurs grandes orientations. L'évaluation produit ainsi une rétroaction périodique concernant les différents aspects du programme (implantation, efficacité, effets non désirés, etc.) utilisée pour effectuer des choix quant à l'avenir de ces programmes.

1.2 La controverse qui existe au sein de l'évaluation de programme

Même si l'évaluation de programme a acquis ses lettres de noblesse grâce à de nombreux apports lui étant spécifiques, il semble se dégager une évidence selon laquelle elle est confrontée, et ce, depuis son origine, à un paradoxe indéniable. En effet, alors que cette activité est en constante demande compte tenu de sa pertinence pour la prise de décision, il n'en demeure pas moins qu'elle est périodiquement remise en question par les utilisateurs concernant sa capacité à remplir ses fonctions premières, soit produire des jugements

suffisamment crédibles pour être acceptés des clients et ainsi, guider les grandes orientations des programmes.

Ainsi, suite à une étude des évaluations de programme en Colombie Britannique, McDavid (2001) conclut que l'évaluation de programme n'arrive toujours pas à émettre des jugements suffisamment fondés pour éclairer la prise de décision. À cet effet, l'étude du Secrétariat du Conseil du Trésor du Canada (2004) semble confirmer ces dires. Cette étude, qui porte sur l'analyse de rapports d'évaluation (115) commandés par différents ministères, révèle que 32% des rapports présentent un raisonnement inadéquat pour appuyer leur jugement. L'étude de Toulemonde (2005), sur l'utilisation des évaluations de programme dans la région du Limousin en France, l'amène également à stipuler qu'il y a relativement peu de clients qui utilisent les jugements qu'elle produit pour orienter l'avenir de leur programme. Finalement, selon les analyses de Virtanen et Uusikylä (2002) qui ont réfléchi sur la complexité des programmes sociaux, la pratique actuelle de l'évaluation de programme ne permet pas, pour le moment, de poser un jugement qui prend en considération toute la complexité des programmes.

Il est également possible de percevoir ce problème à la lecture de nombreux écrits dans le domaine qui portent sur l'avenir qui lui est réservé (Dubois et Marceau, 2005; King, 2003; Patton, 2001). Les nombreuses et régulières remises en question de l'évaluation de programme préoccupent ses auteurs et amènent d'autres à tenter d'expliquer ce phénomène.

À ce sujet, Guba (1972) a été l'un des premiers auteurs à identifier et à documenter ce malaise récurrent. Dans son article intitulé *The Failure of Educational Evaluation*, il émet l'hypothèse que cette situation serait attribuable à la difficulté éprouvée par les évaluateurs à bien cerner la spécificité de la démarche d'évaluation de programme. Ainsi, les évaluateurs comprendraient mal le processus permettant d'identifier les critères qui devrait être logiquement lié au contexte du programme et à celui de son évaluation. Scriven (1995) souscrit à cette analyse et renchérit en avançant que l'évaluation de programme fait face à un problème structurel concernant la logique qui sous-tend la construction du jugement et la formulation de recommandations. Il établit également les éléments essentiels à l'évaluation

qu'il précise à l'intérieur d'une conception de l'évaluation qu'il baptise la *logique de l'évaluation*.

Ainsi, Guba (1972) et Scriven (1995) introduisent qu'il y aurait un processus spécifique à l'évaluation et que ce serait une mauvaise application de ce dernier qui pourrait nuire à la formulation d'un jugement fondé. Cependant, leurs propos constituent essentiellement des hypothèses d'explication.

1.3 Problématique de la recherche

Les propos de Guba (1972) et Scriven (1995) demeurent de l'ordre des réflexions et ils n'ont ainsi jamais fait l'objet d'une vérification scientifique. Certains auteurs ont manifesté de l'intérêt dans ce sens.

À ce sujet, Arens (2006) a récemment étudié la construction des jugements des évaluations identifiées comme exemplaires selon l'*American Évaluation Association* (5). Les résultats indiquent que la façon dont les évaluateurs parviennent à leurs jugements est peu documentée et ne suit pas toujours un parcours logique, ce qui nuirait à la formulation d'un jugement fondé. À ce sujet, l'absence de standard semble également avoir un impact négatif sur cette dernière. L'étude de Hurteau et Houle (2006), qui constitue le premier volet du programme de recherche, a également approfondi cette relation. Ayant élaboré une modélisation du processus spécifique à l'évaluation de programme, ces chercheurs ont analysé si elle se retrouvait à l'intérieur de 69 évaluations de programme ayant fait l'objet d'une publication dans trois revues scientifiques spécialisées (*Journal Canadien en Évaluation de Programme, American Journal of Evaluation, Evaluation and Program Planning*). Les résultats de l'étude mettent en évidence que les éléments de leur modélisation sont présents dans plus de 80% des cas, mais que les jugements ne s'appuient pas sur des standards dans 40% des cas. S'il ressort que cela semble avoir des conséquences négatives sur le jugement, ce constat demeure cependant au niveau des impressions.

C'est ainsi que, chacune à leur façon, les deux études mettent en évidence la difficulté qu'éprouvent les évaluateurs à présenter un jugement suffisamment fondé pour que les clients les acceptent et ainsi, guider les grandes orientations de leur programme. De plus, elles font ressortir l'absence de standards, une composante essentielle au processus, comme étant une cause possible de ce malaise. Il demeure toutefois que les résultats ne permettent pas, pour le moment, de confirmer la relation entre l'application du processus spécifique à l'évaluation de programme et la formulation d'un jugement fondé.

Enfin, il est à noter que l'absence d'une représentation consensuelle du processus spécifique à l'évaluation de programme ajouté au peu d'écrits portant sur ce qui assure de bons fondements au jugement en évaluation pourrait également expliquer pourquoi les hypothèses de Guba (1972) et de Scriven (1995), introduites il y a plus de dix ans, ne sont toujours pas validées scientifiquement à ce jour.

1.4 La question de recherche

Les résultats d'études antérieures laissant ainsi croire que:

- 1) la pratique actuelle en évaluation de programme éprouve de la difficulté à générer un jugement fondé;
- 2) et que cela pourrait être dû à une mauvaise compréhension/application du processus spécifique à l'évaluation de programme;

le second volet du présent programme vise à apporter des éléments de réponse à la question suivante : Dans quelle mesure l'application du processus spécifique à l'évaluation de programme contribue-t-elle à formuler un jugement fondé?

1.5 La pertinence de l'étude

1.5.1 La pertinence scientifique

En ce qui concerne la pertinence scientifique de l'étude, plusieurs auteurs, dont Alkin (2003), Christie (2003b) et King (2003) croient qu'il faut désormais se pencher sur les fondements de l'évaluation de programme en soulignant qu'ils ont été négligés; le peu de développement théorique la concernant portant plutôt sur son instrumentation. Selon cette perspective, l'essence de cette recherche, soit le processus spécifique à l'évaluation de programme, constitue un élément qu'il est plus que pertinent d'approfondir.

Aussi, tel qu'expliqué par King (2003), le manque d'études concernant l'évaluation de programme est flagrant. Plus précisément, le développement théorique de l'évaluation demeure limité pour plusieurs raisons dont:

- 1) un manque de consensus conceptuel et de recherche sur le sujet;
- 2) l'engouement pour d'autres sujets, dont le développement des programmes, de la méthodologie et des modèles d'évaluation;
- 3) et le caractère pratique de ce domaine relativement récent qui ne lui confère pas un appui théorique invitant pour la recherche.

Selon ces propos, cette recherche dépasse plusieurs limites pour s'intéresser à un domaine encore peu exploré scientifiquement et se situe au cœur d'un enjeu important pour le développement global du domaine de l'évaluation de programme, soit le processus qui lui est spécifique.

1.5.2 La pertinence sociale

Au cours des dernières décennies, l'évaluation de programme s'est vu confier un rôle d'une grande importance, c'est-à-dire guider les grandes orientations de la majorité des programmes éducatifs et sociaux de nos sociétés. Conséquemment, ce domaine se doit de posséder un raisonnement assez rigoureux pour assurer la construction d'un jugement juste et fiable pour les sociétés qui en seront affectées. Puisqu'elle vise l'amélioration des pratiques évaluatives à l'intérieur de ces domaines, la présente étude gagne une grande pertinence.

Plus spécifiquement, la pertinence sociale de cette recherche trouve également une place de choix au sein de l'actualité québécoise. À ce sujet, une récente réforme des programmes scolaires a stimulé l'intérêt de cette population pour les résultats de l'évaluation de ces programmes. Les jugements rendus lors de leur évaluation ont donc tout avantage à être solidement fondés pour assurer leur crédibilité aux yeux des Québécois. Selon cette même perspective, cette population a un intérêt grandissant concernant les résultats des évaluations des programmes relatifs à l'éducation en tant qu'outil d'amélioration des conditions sociales. Cela concerne notamment les programmes visant à prévenir la toxicomanie, la violence, la délinquance, le tabagisme, le stress ou la détresse émotionnelle.

1.6. Synthèse

Pour conclure, l'apport du processus spécifique à l'évaluation de programme doit être sérieusement pris en considération, car il pourrait constituer une éventuelle solution au malaise qu'éprouve l'évaluation de programme concernant sa capacité à produire des jugements suffisamment fondés pour être acceptés des clients et guider les grandes orientations d'un programme. D'un point de vue scientifique, la pertinence de cette recherche s'appuie sur le fait qu'elle approfondit deux éléments théoriques essentiels au maintien et au développement de ce domaine, soit le processus qui lui est spécifique et la formulation de jugements fondés. D'un point de vue social, la pertinence de cette recherche s'appuie sur le fait qu'elle pourrait contribuer à améliorer la qualité des pratiques évaluatives liées aux évaluations des grands programmes éducatifs de notre société. Enfin, puisque cette recherche

se penche sur les éléments qui encadrent la pratique, elle constitue d'autant plus une approche intéressante à une ère où tout doit être fait rapidement pour répondre aux exigences des clients et malheureusement, parfois aux dépens de la qualité du raisonnement employé.

CHAPITRE II

CADRE THÉORIQUE

Ce deuxième chapitre s'intéresse aux fondements théoriques des concepts abordés à l'intérieur de la problématique. À cet effet, la distinction entre les théories descriptives et normatives y est établie suivie de l'approfondissement des théories descriptives ayant contribué au développement du processus spécifique à l'évaluation de programme (Scriven, 1980; Hurteau, 1991; Fournier, 1995; Stake, 2004 et Arens, 2006). Les détails de la modélisation de ce processus proposée par Hurteau et Houle (2006) y sont ensuite présentés. Enfin, il se termine par la présentation des objectifs de recherche découlant de la question de recherche énoncée au chapitre précédent.

2.1 Distinction entre les théories descriptives et normatives

Le développement de l'évaluation de programme s'est effectué grâce à l'introduction de théories normatives et descriptives. Si les théories normatives dictent aux évaluateurs la manière dont ils devraient effectuer leurs évaluations, les théories descriptives tentent, au contraire, de comprendre la nature de ses différentes opérations ainsi que celle de leurs interrelations. (Christie 2003a). Le tableau 2.1 en rassemble les principaux exemples et établit leur lien avec le processus spécifique à l'évaluation de programme.

Tableau 2.1
Distinction entre les théories descriptives et normatives

Nature des théories	Descriptives	Normatives
Exemples	<p><i>La logique de l'évaluation</i> (Scriven, 1980)</p> <p><i>Les choix stratégiques</i> (Hurteau, 1991)</p> <p><i>Les niveaux de logique</i> (Fournier, 1995)</p> <p><i>Les types de pensées évaluatives</i> (Stake, 2004)</p> <p><i>Le raisonnement évaluatif</i> (Arens, 2006)</p>	<p>La théorie soutenant la <i>cueillette et l'analyse des données</i> (Shadish, Cook et Campbell, 2002)</p> <p>La théorie soutenant les <i>modèles d'évaluation</i> (Stufflebeam, 2001)</p> <p>La théorie soutenant les <i>standards de pratique</i> (Sanders et coll., 1994)</p>
Traite	Processus spécifique à l'évaluation de programme	Instrumentation et modalités qui supportent ce processus

Les théories descriptives, puisqu'elles visent à décrire et comprendre la pratique, étudient la nature du geste posé en évaluation de programme. Scriven (1980) est le premier auteur ayant abordé cette distinction qu'il rend explicite par sa conception de l'évaluation qu'il baptise la *logique de l'évaluation* (Mathison, 2004). Plusieurs auteurs ont poursuivi cette réflexion dont Hurteau (1991), Fournier (1995), Stake (2004) et Arens (2006) et l'ont adaptée au contexte d'une évaluation de programme. Hurteau et Houle (2006) ont synthétisé ces différentes théories descriptives pour créer une modélisation du processus spécifique à l'évaluation de programme, présentée en détail à la section suivante. Ainsi, le processus spécifique à l'évaluation de programme demeure hors de l'influence des théories normatives s'appuyant plutôt sur des théories descriptives.

Les théories normatives portent sur l'instrumentation et les modalités qui supportent ce processus puisqu'elles visent, au contraire, à établir ce que les évaluateurs doivent faire. Elles peuvent toucher les théories soutenant la cueillette et l'analyse des données, les modèles d'évaluation (traduisant la vision d'un auteur sur le rôle d'un évaluateur et sa place respective au sein de la démarche) ou encore les standards de pratique (assurant une évaluation de qualité selon un commun accord sur le sujet concernant l'utilité, la faisabilité, l'éthique ou la validité de toute évaluation de programme). Au regard des différentes publications à l'intérieur de ce domaine, la majorité des théories en évaluation de programme sont de type normative puisqu'elles touchent à l'instrumentation et les modalités qui supportent le processus spécifique à l'évaluation de programme.

Le tableau 2.1 fait donc ressortir le fait que les théories descriptives traitent essentiellement du processus spécifique à l'évaluation de programme et les théories normatives, au contraire, de l'instrumentation et des modalités supportant ce processus.

Pour conclure, tel qu'abordé à l'intérieur de la problématique, nombreux chercheurs sont d'avis que la recherche doit désormais laisser de côté les théories normatives pour se tourner vers la création de théories descriptives (Alkin, 2004; Christie, 2003b; King, 2003). La présente recherche poursuit cette réflexion en portant exclusivement sur les opérations qui caractérisent l'évaluation de programme. Ainsi, le cadre théorique de cette recherche s'appuie essentiellement sur des théories descriptives et concerne exclusivement le processus spécifique à l'évaluation de programme qui sera maintenant décrit.

2.2 Le processus spécifique à l'évaluation de programme

Cette section présente les théories descriptives ayant contribué au développement du processus spécifique à l'évaluation de programme. À cet effet, la *logique de l'évaluation* introduite par Scriven (1980) y sera premièrement présentée suivie des *choix stratégiques et méthodologiques* de Hurteau (1991), de la *logique de travail* de Fournier (1995) et de la *pensée interprétative* de Stake (2004). Viendront enrichir ces apports théoriques, les propos

d'Arens (2006) concernant la pondération des critères et la synthèse de ces contributions effectuée par Hurteau et Houle (2006) dont les détails seront fournis à la section suivante.

2.2.1 La contribution de Scriven (1980): la logique de l'évaluation

Scriven (1980) contribue au développement du processus spécifique à l'évaluation de programme en étant le premier à cerner les éléments ainsi que des composantes qui caractérisent l'acte même d'évaluer. Il le fait en se référant aux principes de raisonnement propre à la logique générale qui relèvent de la déduction, de l'induction et de l'inférence statistique. Il les approfondit à l'intérieur de sa conception de l'évaluation qu'il baptise la *logique de l'évaluation*.

Le premier élément consiste à établir *les critères*, c'est-à-dire les éléments de l'objet qui seront évalués. Ils sont définis comme les aspects, les attributs ou les dimensions qui caractériseront la valeur ou le mérite de l'objet d'évaluation.

Le deuxième élément consiste à fixer les *standards*, c'est-à-dire le ou les points de coupure qui déterminent le niveau de performance à atteindre pour que le jugement soit favorable. Ils constituent un seuil, typiquement exprimé en pourcentage ou en degré de qualité, qui permettra d'établir si le jugement sera positif ou négatif.

Le troisième élément consiste à *mesurer la performance et la mettre en relation avec les standards*. Cette activité implique une cueillette de données, mais ne la concerne pas directement, car l'essence de cet élément est la mise en relation des données obtenues avec les standards fixés.

Le quatrième élément consiste à *synthétiser et intégrer les informations pour produire le jugement*.

Enfin, par l'introduction de la *logique de l'évaluation*, Scriven introduit les bases théoriques du processus spécifique à l'évaluation de programme. Ces éléments, lorsque suivis logiquement, constituent l'acte d'évaluer, c'est-à-dire celui d'accorder une valeur soit par l'attribution d'une note, d'un rang, d'un résultat ou encore d'un indicateur de proportion à l'objet de l'évaluation.

2.2.2 La contribution de Hurteau (1991): les choix stratégiques et méthodologiques

Hurteau (1991) contribue également au développement du processus spécifique à l'évaluation de programme en avançant que les *choix méthodologiques*, c'est-à-dire ceux qui concernent la méthode de cueillette et d'analyse des données, ne font pas partie du processus spécifique à l'évaluation de programme. Au contraire des *choix stratégiques* référant au choix de l'objet de l'évaluation, à celui des critères et des standards (qui reprennent également les éléments de la *logique de l'évaluation* (Scriven, 1980), les *choix méthodologiques* seraient exclus de ce processus.

En effet, même si les *choix méthodologiques* occupent une grande place à l'intérieur des écrits du domaine de l'évaluation, le lecteur se rappellera qu'ils appartiennent à plutôt à l'instrumentation issue de théories normatives et non de théories descriptives dont le processus spécifique à l'évaluation de programme fait partie. Les forces et des faiblesses de chacune des méthodes pour recueillir et analyser les données sont des considérations multidisciplinaires et viennent plutôt soutenir le processus spécifique à l'évaluation de programme sans toutefois en faire partie. En d'autres mots, même s'il est inévitablement d'effectuer des *choix méthodologiques* lors d'une évaluation de programme, ils ne constituent pas une opération concernant assez directement le processus permettant de construire un jugement pour constituer une composante lui étant spécifique.

Aussi, cette auteure met en évidence que les *choix méthodologiques* découlent des *choix stratégiques*. Selon cette hypothèse, une plus grande cohérence serait assurée en établissant

les *choix stratégiques*, encadrant les orientations majeures de la démarche, avant les *choix méthodologiques*, précisant la manière dont les données seront recueillies et traitées.

Enfin, par l'établissement de la distinction entre les *choix stratégiques* et *méthodologiques*, Hurteau (1991) précise que le processus spécifique ne contient pas les *choix méthodologiques*. Elle démontre également qu'une plus grande cohérence serait assurée en établissant les *choix stratégiques*, encadrant les orientations majeures de la démarche avant les *choix méthodologiques*.

2.2.3 La contribution de Fournier (1995): la logique de travail

Fournier (1995) a également contribué au développement du processus spécifique à l'évaluation de programme par l'ajout de nouveaux éléments, regroupés sous le nom de *logique de travail*, qui, à l'opposé de la *logique générale* ayant inspiré la *logique de l'évaluation* (Scriven, 1980), serait spécifique à l'évaluation de programme. Ces éléments seraient utilisés en alternance avec ceux introduits par Scriven (1980).

Le premier élément de la *logique de travail* consiste à déterminer les besoins des gestionnaires qui recevront l'évaluation.

Le deuxième élément consiste à cerner avec précision l'objet à l'étude selon les besoins identifiés.

Le troisième élément consiste à formuler une question d'évaluation qui prend en considération tous ces éléments.

Enfin, par l'introduction de la *logique de travail*, Fournier (1995) enrichit le processus spécifique à l'évaluation de programme en y incluant des éléments plus pratiques et opérationnels en contexte d'évaluation de programme.

2.2.4 La contribution de Stake (2004): la pensée interprétative

Stake (2004) a lui aussi contribué au développement du processus spécifique à l'évaluation de programme en spécifiant qu'il implique deux types de pensées évaluatives utilisées simultanément, soit la *pensée critériée* et la *pensée interprétative*.

La *pensée critériée* est celle qui est liée à la rationalité, la rigueur, au raisonnement logique et au sens objectif de l'évaluateur. Ce type de pensée permet d'effectuer la mesure des variables à l'étude le plus précisément possible et d'arriver aux conclusions les plus crédibles possible. Elle comprend l'identification de critères et leur mise en relation avec des standards pour porter un jugement, soit l'ensemble des éléments de la *logique de l'évaluation* (Scriven, 1980).

En ce qui concerne la *pensée interprétative*, elle fait, au contraire de la pensée critériée, référence à la subjectivité que possède inévitablement l'évaluateur, les clients de l'évaluation ainsi que les personnes impliquées au sein du programme. L'évaluateur l'utilisera lorsqu'il devra traduire sa perception de la « qualité » relative à l'objet d'évaluation, en critères et en standards. De plus, c'est sa connaissance du contexte du programme et celui de son évaluation (budget, temps, nature de l'objet de l'évaluation, relations avec le client, accès à l'information, etc.) qui constituera le plus grand appui pour justifier cette subjectivité.

Enfin, par l'introduction de la théorie liée à la *pensée interprétative*, Stake (2004) enrichit le processus spécifique à l'évaluation de programme en lui fournissant un lien avec la subjectivité de l'évaluateur qui s'appuie sur sa connaissance du contexte du programme et de son évaluation.

2.2.5 La contribution d'Arens (2006) : la pondération des critères

Arens (2006) a également contribué au développement du processus spécifique à l'évaluation de programme. Ses travaux s'inspirent de ceux de Toulmin (1993) et de Fournier (1995). Elle propose une représentation de la démarche de construction des jugements qui s'appuie sur le rôle de l'*inférence* et des *évidences* au sein de la construction des jugements (voir figure 2.1).

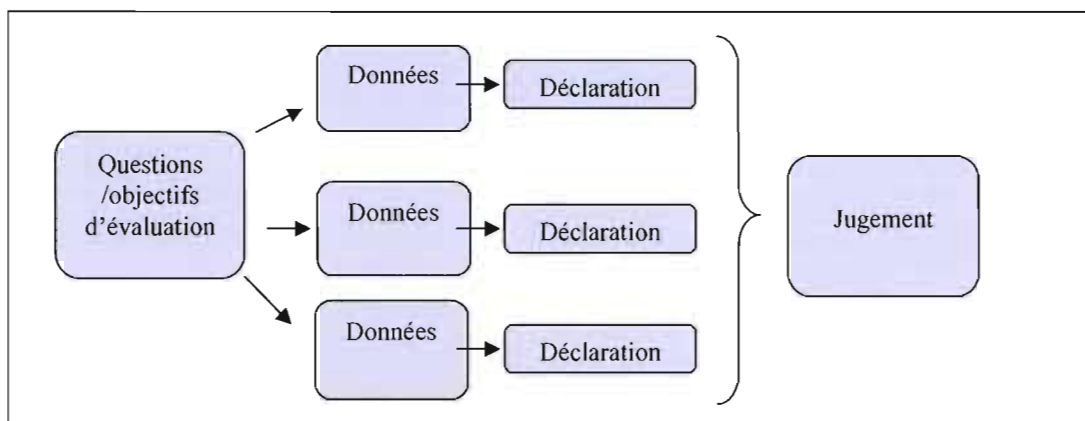


Figure 2.1 Le jugement: une synthèse pondérée de différentes déclarations selon Arens (2006).

Selon cette auteure, construire un jugement consiste à recueillir des évidences empiriques (information pertinente) qui seront synthétisées en déclarations afin de produire un jugement. Alors que la déclaration s'appuie, règle générale, sur les critères (exemple: 80% des clients sont satisfaits), le jugement reprend cette information en appliquant le standard (exemple: le programme est efficace parce qu'il se conforme au standard lié au critère de satisfaction des clients qui avait été établi à 75%). Les différentes déclarations sont ensuite synthétisées pour produire un jugement global sur l'aspect du programme à l'évaluation. Cette synthèse peut être effectuée par diverses méthodes qui ajoutent une sorte de « pondération » aux différentes déclarations (House, 1995; House et Howe, 1999; Scriven,

1991; What Works Clearinghouse, 2003 dans Arens, 2006). Cette représentation spécifie que le jugement constitue une synthèse pondérée de différentes déclarations, que cette synthèse soit rendue explicite ou non à l'intérieur du rapport. Selon cette auteure, si la plupart des évaluateurs ne rendent pas explicite le fait qu'ils aient utilisé une méthode ou l'autre, ils en utilisent toujours toutefois une, même inconsciemment, pour synthétiser les différentes déclarations produites en un jugement.

Enfin, par son approfondissement théorique de l'opération consistant à pondérer les différentes déclarations pour les synthétiser en un jugement, Arens (2006) avance que les critères sont pondérés à l'intérieur du processus spécifique à l'évaluation de programme.

2.2.6 La contribution de Hurteau et Houle (2006) : une modélisation

La contribution de Hurteau et Houle (2006) au processus spécifique à l'évaluation de programme est considérable, car ils ont synthétisé l'ensemble de ces différentes contributions précédentes (Scriven, 1980; Hurteau, 1991; Fournier, 1995; Stake, 2004 et Arens, 2006), selon la méthodologie propre à l'anasynthèse, pour produire une modélisation du processus spécifique à l'évaluation de programme. Cette modélisation se retrouve à la figure 2.2 accompagnée des fondements théoriques sur lesquels s'appuie chacune de ses composantes. Il est à noter que l'ensemble des auteurs consultés reconnaissent que les *choix stratégiques* constituent le cœur de l'évaluation de programme, et ce, même s'ils n'y réfèrent pas selon ces termes et que les choix *méthodologiques* relèvent de considérations différentes (Hurteau, 1991). Cette modélisation a fait l'objet d'une validation au premier volet du projet de recherche. À cet effet, une analyse du contenu de 69 rapports d'évaluation publiés dans des revues spécialisées a démontré qu'elle représente réellement ce qui se fait dans les pratiques exemplaires bien que certains éléments, dont l'utilisation de standards, soient moins présents.

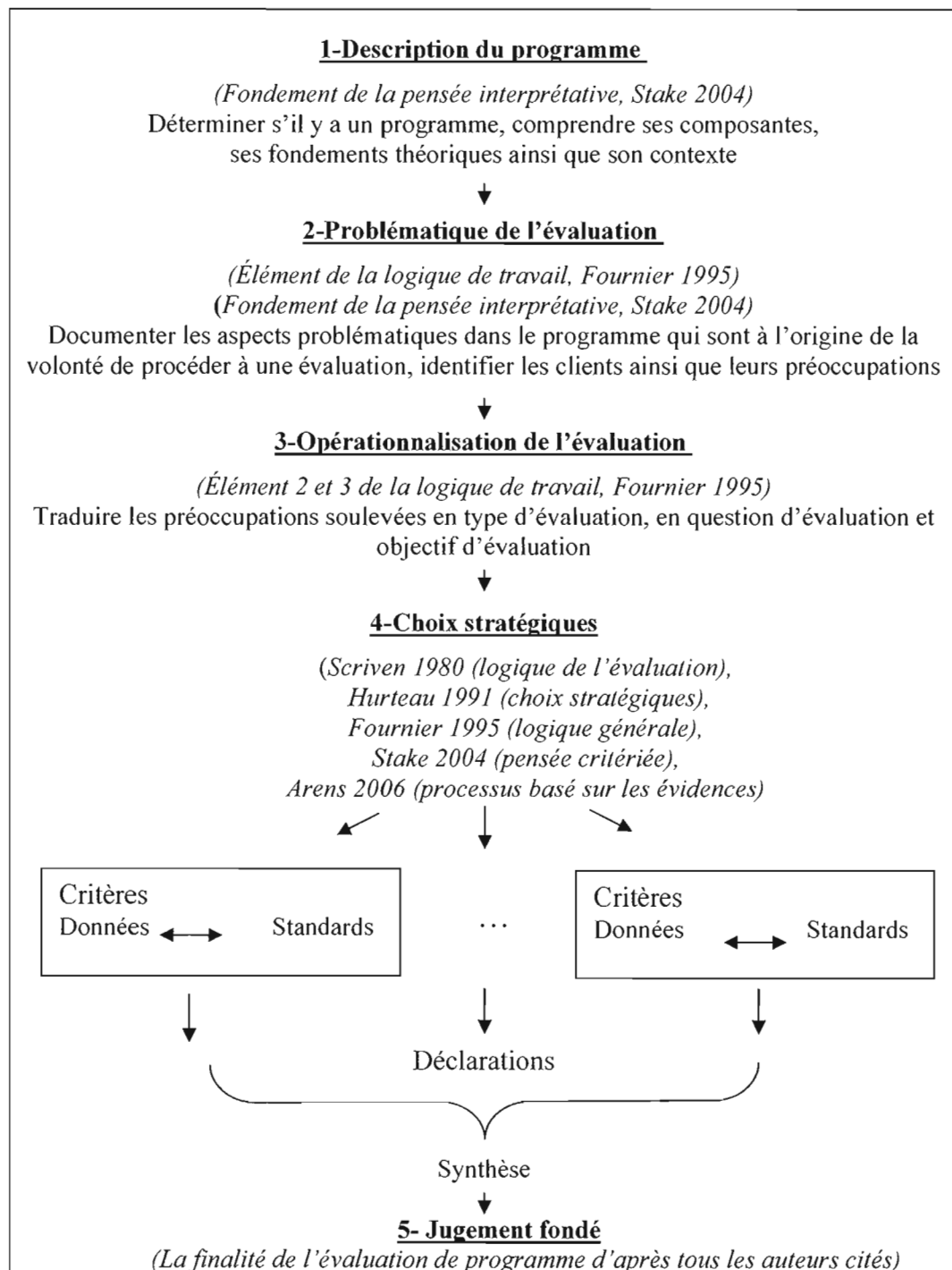


Figure 2.2 La modélisation du processus spécifique à l'évaluation de programme et les fondements théoriques de ses composantes. (Hurteau et Houle, 2006)

2.3 Les constituants de la modélisation

Bien que les principaux fondements théoriques du processus spécifique à l'évaluation de programme aient été abordés à la section précédente (Scriven, 1980; Hurteau, 1991; Fournier, 1995, Stake, 2004; Arens, 2006), les auteurs de la modélisation (Hurteau et Houle, 2006) ont utilisé d'autres contributions à l'intérieur de leur synthèse pour compléter chacune des composantes. C'est pourquoi cette section reprend chacune des composantes de la modélisation et les approfondit selon ces différents apports théoriques.

2.3.1 La description du programme

La première composante, intitulée simplement *Description du programme*, réfère aux aspects qu'il est nécessaire de documenter afin de s'assurer qu'il y a bien un programme qui pourra éventuellement faire l'objet d'une évaluation. À cet effet, l'apport de Chen (2005) qui porte sur ce qui constitue un programme est abordé afin de préciser les éléments de cette composante.

Chen (2005) définit un programme comme un système ouvert et systémique, composé d'un ensemble de ressources (humaines, matérielles et financières) et de moyens (interventions, activités) appelés *intrants*. Il est mis en œuvre à l'intérieur d'un *processus* en interaction avec son environnement pour atteindre des objectifs liés de près ou de loin à l'amélioration du bien-être de sa clientèle dont les *résultats* rendent compte du degré de transformation atteint. La figure 2.3 illustre ces propos en représentant le programme un peu comme un dispositif transformant les intrants en résultats.

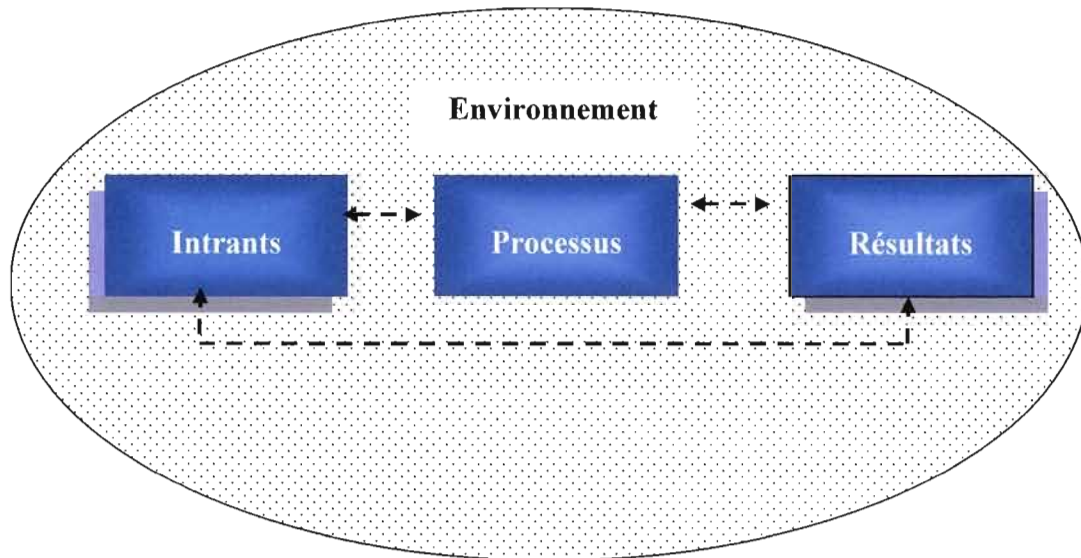


Figure 2.3 Un programme: un système. (Tirée de Chen, 2005)

Les *Intrants* incluent tous les éléments en place au moment de la conception du programme qui pourront être mobilisés pour élaborer le programme (les ressources financières, humaines, matérielles, etc.)

Le *Processus*, quant à lui, inclut l'ensemble des opérations qui ont servi à organiser l'intervention (formation des intervenants, élaboration d'un calendrier des activités, etc.) et à l'intervention elle-même (la durée de l'intervention, son contenu, sa fréquence, ses caractéristiques, etc.)

L'élément *Résultats* réfère à ce qui est attendu comme changement chez la clientèle cible (apprentissage d'une nouvelle langue, diminution d'actes de violence chez un individu, etc.) Ils peuvent être à courts, moyens et longs termes.

Ce modèle s'avère un outil de choix pour confirmer l'existence d'un programme qui pourra faire, éventuellement, l'objet d'une évaluation. Plus précisément, certaines informations se révèlent essentielles à la compréhension globale d'un programme dont la clientèle à qui s'adresse le programme, les caractéristiques de l'intervention (élément du

Processus), le cadre théorique sur lequel elle s'appuie (éléments des *Intrants*) et les objectifs du programme (formulation des *Résultats* souhaités).

Enfin, selon ces considérations, effectuer les différentes opérations liées à la première composante de la modélisation, soit la *Description du programme*, consiste à saisir la clientèle à qui s'adresse le programme, les objectifs du programme, les caractéristiques de l'intervention et le cadre théorique sur lequel il s'appuie.

2.3.2 La problématique de l'évaluation

La deuxième composante, intitulée *Problématique de l'évaluation*, constitue une étape cruciale à la démarche puisqu'elle inclut l'élément qui a déclenché l'évaluation (politique d'évaluation, problème rencontré, besoin d'information, etc.) et les préoccupations des gestionnaires qui ont demandé l'évaluation. À cet effet, les apports de Scriven (1980) et Fournier (1995) qui approfondissent les préoccupations des clients de l'évaluation seront abordés afin de préciser les éléments de cette composante.

Les préoccupations des clients de l'évaluation touchent principalement deux aspects, soit le type de jugement (Scriven, 1980), et le type d'argumentation qu'ils attendent (Fournier, 1995). Le type de jugement attendu peut être, par exemple, l'attribution d'un rang à l'objet de l'évaluation suite à une classification ou d'une classe suite à une comparaison. Quant au type d'argumentation attendu, elle peut être fondée sur l'opinion d'experts sur le sujet, sur des résultats d'un schéma expérimental ou encore sur l'opinion de participants. Étant donné que les conclusions produites par l'évaluation ont pour objectif de répondre à cette problématique, l'évaluateur se basera sur ces différentes considérations, au moment opportun, pour orienter ses choix (Fournier, 1995).

Enfin, selon ces considérations, effectuer la deuxième composante de la modélisation, soit la *Problématique de l'évaluation*, consiste à comprendre l'élément qui a déclenché

l'évaluation ainsi que les attentes des clients de l'évaluation concernant le type de jugement et d'argumentation escomptés.

2.3.3 L'opérationnalisation de l'évaluation

La troisième composante, intitulée *Opérationnalisation de l'évaluation*, permet de traduire les besoins présents à l'intérieur de la problématique en langage appartenant au domaine de l'évaluation de programme pour structurer la démarche, soit les questions/objectifs et l'objet d'évaluation. À cet effet, les apports de Fournier (1995) et Rossi, Lipsey et Freeman (2004) qui portent sur les liens qui unissent ces éléments sont abordés afin de préciser les éléments de cette composante.

Premièrement, Fournier (1995) stipule que l'objet de l'évaluation sera influencé par le type de jugement attendu tel que retrouvé à l'intérieur de la *Problématique de l'évaluation*. Deuxièmement, Rossi, Lipsey et Freeman (2004) confirment cette relation et précisent que ces choix influenceront également, à leur tour, l'ensemble des composantes qui suivront. De plus, ils introduisent une typologie qui établit le lien qui unit les types d'évaluation aux questions (et/ou d'objectifs selon le choix de l'évaluateur) présentée au tableau 2.2

Tableau 2.2
Relation entre les questions (et/ou objectifs) et le type d'évaluation
selon Rossi, Lipsey et Freeman (2004)

Type d'évaluation	Exemple de questions et/ou d'objectifs reliés
Évaluation de pertinence (Needs assessment)	Question : Quels sont les besoins de la population cible? Objectif : Établir les besoins de la population cible
Évaluation de la conception ou de la planification (Assessment of program theory)	Question : Quel service répondrait au besoin identifié? Objectif : Établir quel service répondrait au besoin identifié
Évaluation de l'implantation (Assessment of program process)	Question : Est-ce que le programme est implanté tel que prévu? Objectif : Établir si le programme est implanté tel que prévu.
Évaluation de l'impact (Impact assessment)	Question : Est-ce que le programme atteint ses objectifs? Objectif : Établir si le programme atteint ses objectifs
Évaluation de l'efficacité ou du coût-bénéfice (Efficiency assessment)	Question : Est-ce que le programme pourrait atteindre les objectifs à moindre coût? Objectif : Établir si le programme pourrait atteindre les objectifs à moindre coût

Enfin, selon ces considérations, effectuer les différentes opérations liées à la troisième composante de la modélisation, soit *l'Opérationnalisation de l'évaluation* consiste à traduire les besoins présents à l'intérieur de la problématique en objet d'évaluation et en questions/objectifs et ce de façon cohérente, c'est-à-dire en respectant minimalement la typologie de Rossi, Lipsey et Freeman (2004).

2.3.4 Les choix stratégiques de l'évaluation

La quatrième composante, intitulée *Choix stratégiques*, consiste essentiellement à identifier les critères et les standards sur lesquels reposera le jugement. Ils constituent la pierre angulaire de la démarche évaluative puisque ce sont de ces choix que découlera la cueillette de données sur lesquelles s'appuiera le jugement (Hurteau, 1991). À cet effet, les

apports d'Arens (2006), de Hurteau (1991), de Fink (2004) et de Fournier (1995) qui approfondissent ce qui constitue les *Choix stratégiques* sont abordés afin de préciser les éléments de cette composante.

Les critères constituent les attributs du programme choisis pour leur bonne représentativité de l'aspect du programme à l'évaluation selon les différentes perceptions des clients de leur programme et d'après leurs attentes quant aux objectifs de l'évaluation. Tel que mentionné à l'intérieur de la contribution de Arens (2006) à la modélisation, même si les évaluateurs ne rendent pas explicite le fait qu'ils aient utilisé une méthode pour attribuer aux critères une sorte de pondération, ils en utilisent toujours une même inconsciemment, afin de synthétiser les différentes déclarations y étant liées en un jugement. Plusieurs méthodes sont proposées (House, 1995; House et Howe, 1999; Scriven, 1991; What Works Clearinghouse, 2003 dans Arens 2006) même si elles ne semblent pas être utilisées en pratique.

Les standards, quant à eux, fixent le point de coupure entre un jugement favorable et défavorable concernant chacun des critères identifiés (ex. norme nationale, groupe contrôle, standards préétablis selon les valeurs des gestionnaires qui ont demandé l'évaluation,...). Ils peuvent être identifiés en ayant recours à diverses méthodes, dont une comparaison avec d'autres programmes, une comparaison avec les données des années précédentes du programme, selon l'avis d'un ou des experts du domaine, selon une étude des écrits sur le sujet ou encore selon des données publiques (Fink, 2004) en tenant compte du type d'argumentation attendu par les clients de l'évaluation (Fournier, 1995).

Aussi, tel que mentionné par Hurteau (1991), le moment où sont introduits les critères et les standards est un élément déterminant concernant la cohérence de la démarche. En effet, le choix des instruments de mesure et d'analyse, c'est-à-dire le détail des *choix méthodologiques*, ne peut être établi sans avoir préalablement sélectionné le choix des critères et des standards, les *choix stratégiques*. Une plus grande cohérence serait ainsi assurée en précisant les *choix stratégiques* encadrant les orientations majeures de la démarche, avant les *choix méthodologiques* qui en découlent.

Enfin, selon ces considérations, effectuer la quatrième composante de la modélisation, soit *les Choix stratégiques*, consiste à spécifier avant la cueillette de données, les critères et les standards sur lesquels le jugement se fondera en tenant compte du type d'argumentation attendu par les clients de l'évaluation en plus d'annoncer leur pondération.

2.4 Le jugement fondé

La pratique évaluative possède son propre raisonnement (Taylor, 1973; Toulmin, 1993; Toulmin, Rieke et Janik, 1984; McCarthy, 1973; Redding, 1989). Il est constitué de deux opérations à la fois distinctes, complémentaires et reliées, soit le développement d'un jugement et son argumentation. Arens (2005, 2006), dont les travaux s'inspirent de ceux de Toulmin (1993) et de Fournier (1995), documente le développement d'un jugement (*voir* section 2.2.5). Quant à la seconde opération, soit l'argumentation, elle bonifie la première dans le but de produire des déclarations qui prennent la forme d'un exposé qui se veut vrai (Trelogan, 2001).

Les objectifs de cette étude se situant davantage au niveau des fondements du jugement que de son argumentation comme telle, des indicateurs témoignant d'un jugement fondé ont été identifiés. Les constats qui sont ressortis de l'étude d'Arens (2006) peuvent suggérer des indicateurs témoignant d'un jugement fondé. Selon une perspective plus philosophique, le modèle de Toulmin (1993) concernant les jugements, peut également suggérer d'autres indicateurs témoignant d'un jugement fondé. Ils seront approfondis à cette section pour préciser ce que nous entendons par « jugement fondé » à l'intérieur de cette recherche.

En ce qui a trait à la contribution d'Arens (2006), en plus d'avoir élaboré une représentation de la construction d'un jugement, elle a effectué plusieurs constats en examinant les jugements retrouvés à l'intérieur des rapports d'évaluations jugés exemplaires (5) par l'American Evaluation Association.

Le premier constat d'Arens (2006) signale que les évaluateurs ne documentent pas la façon dont les différentes déclarations sont synthétisées pour produire le jugement, ce qui nuirait au jugement. En ce sens, un indicateur témoignant d'un jugement fondé pourrait être le fait de rendre explicite la manière dont les différentes déclarations ont été synthétisées pour produire un jugement.

Son deuxième constat indique que plusieurs des questions et objectifs à l'origine de la démarche évaluative ne trouvent pas de réponses claires à l'intérieur du rapport d'évaluation, ce qui nuirait au jugement. En ce sens, un indicateur témoignant d'un jugement fondé pourrait être le fait d'apporter des réponses aux questions ou objectifs d'évaluation.

Enfin, son troisième constat montre que plusieurs recommandations, c'est-à-dire les suggestions de modifications à apporter au programme qui élargissent la portée du jugement, ne concernent pas l'aspect du programme à l'évaluation, ce qui nuirait également au jugement. En ce sens, un indicateur témoignant d'un jugement fondé pourrait être le fait d'émettre des recommandations cohérentes avec l'aspect du programme à l'évaluation.

La figure 2.4 illustre ces indicateurs (en blanc), incorporés à sa représentation de la construction d'un jugement (en gris).

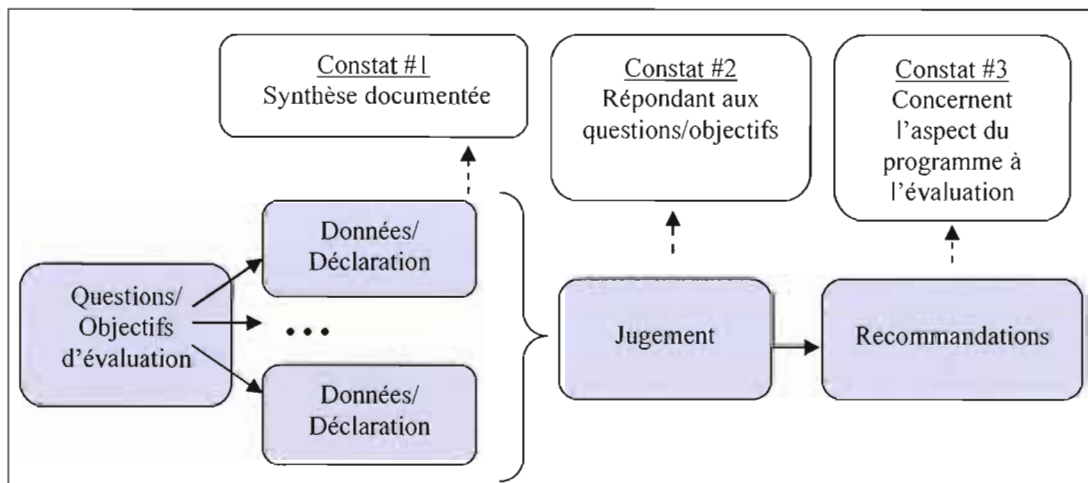


Figure 2.4 Les indicateurs témoignant d'un jugement fondé inspirés des constats d'Arens (2006) (en blanc) incorporés à sa représentation de la construction d'un jugement (en gris).

En ce qui concerne la contribution de Toulmin (1993), il a analysé les jugements selon des contributions d'ordre philosophique. Il énonce qu'un jugement possède de bons fondements dans la mesure où la transformation des données en déclarations, qui pourront éventuellement être transformées en jugement, est justifiée. Plus précisément, cette justification se dira «suffisante» lorsqu'elle permet le passage des données au jugement. Ainsi, un indicateur témoignant d'un jugement fondé pourrait être le fait de justifier les critères et les standards choisis afin de répondre à la question d'évaluation.

Aussi, toujours selon ce modèle, le jugement se retrouve nuancé par diverses considérations qui influencent le degré de certitude ou d'assurance avec lequel il est présenté. Ces considérations font appel aux conditions d'exception du jugement. En ce sens, un indicateur témoignant d'un jugement fondé en évaluation de programme pourrait être le fait de nuancer le jugement.

La figure 2.5 illustre ces indicateurs (en blanc) incorporés à la représentation de la construction d'un jugement d'Arens (2006) (en gris).

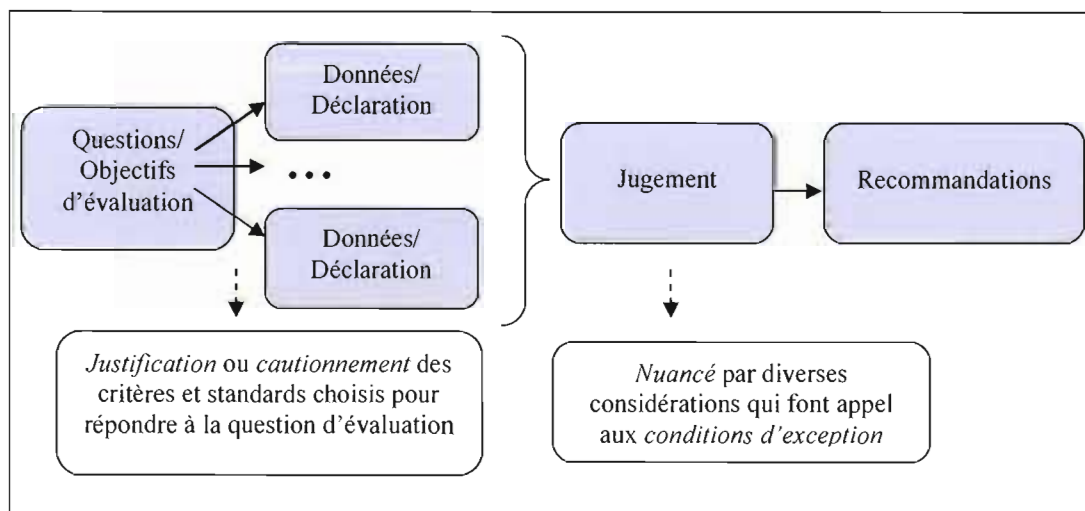


Figure 2.5 Les indicateurs témoignant d'un jugement fondé inspirés de la contribution de Toulmin (1993) (en blanc) incorporés à la représentation de la construction d'un jugement d'Arens (2006) (en gris).

En conclusion, même si Arens (2006) et Toulmin (1993) traitent de la qualité des jugements au sein de leurs travaux, les indicateurs qui en ont été inspirés ne prétendent pas le faire. Ils forment plutôt des indicateurs servant à caractériser les jugements ainsi que leurs fondements. Même si ces contributions ne se retrouvent plus dans le même contexte que celui de leur création, leur intérêt n'est pas négligeable car ils appuient les assises théoriques de cette recherche.

2.5 Synthèse et objectifs de recherche

Au regard des éléments approfondis à l'intérieur de cette section, le processus spécifique à l'évaluation de programme comprend de multiples composantes qui sont toutes reliées entre elles par une logique qui leur est propre. Modélisé par Hurteau et Houle (2006), ce processus s'appuie sur l'anasynthèse des différentes contributions théoriques pertinentes (Scriven, 1980; Hurteau, 1991; Fournier, 1995 et Stake, 2004) et d'approfondissement (Chen, 2005; Rossi, Lipsey et Freeman, 2004; Toulmin, 1993; Arens, 2006).

Si plusieurs auteurs ont proposé un processus à l'évaluation tels que Donaldson et Lipsey, à l'intérieur du livre *The Sage Handbook of Evaluation* (Shaw, Greene et Mark 2006), il demeure que ces propositions, qui s'appuient sur une réflexion et non une démarche scientifique, confirment les composantes faisant partie de la modélisation.

Si la problématique de la présente recherche a conduit à l'identification de la question suivante: *Dans quelle mesure l'application du processus spécifique à l'évaluation de programme contribue-t-elle à formuler un jugement fondé?*, le cadre théorique, lui, aura permis de la rendre opérationnelle au moyen des trois objectifs suivants :

- 1- Établir dans quelle mesure la modélisation du processus spécifique à l'évaluation de programme se retrouve au sein des pratiques évaluatives;
- 2- Caractériser les jugements qui y sont formulés;
- 3- Étudier la contribution de la modélisation à la formulation d'un jugement fondé.

CHAPITRE III

LA DÉMARCHE MÉTHODOLOGIQUE

Ce troisième chapitre spécifie comment le plan de recherche permet de cerner la contribution de l'application du processus spécifique à l'évaluation de programme à la formulation d'un jugement fondé. À cet effet, les choix relatifs au type de recherche, à l'échantillon, à la collecte et à l'analyse des données y sont présentés suivis des forces et faiblesses de l'étude.

3.1 Le type de recherche

La présente recherche se situe au confluent de la recherche descriptive et exploratoire. Plus précisément, elle s'inscrit dans une perspective descriptive puisqu'elle décrit la pratique de l'évaluation de programme d'après la modélisation du processus spécifique à l'évaluation de programme. D'un autre point de vue, l'état actuel des travaux ne permettant pas de cerner tous les enjeux liés au jugement, la présente étude laisse aussi une grande place à l'exploration des éléments qui semble faire en sorte que le jugement soit suffisamment fondé pour être accepté.

3.2 Les choix méthodologiques

3.2.1 Le type d'information requise

L'objet de la présente recherche étant les pratiques en évaluation de programme, l'équipe de chercheurs a décidé de maintenir la même orientation que celle du premier volet du programme de recherche, à savoir analyser le contenu des rapports d'évaluation plutôt que, par exemple, procéder à une enquête auprès des praticiens.

Le premier volet du programme de recherche a analysé des rapports d'évaluation ayant fait l'objet d'un article à l'intérieur d'une revue scientifique parce qu'il visait à valider la modélisation auprès des « pratiques exemplaires ». Le présent volet porte plutôt sur de rapports d'évaluation originaux afin de se rapprocher davantage de ce qui constitue la pratique.

3.2.2 Le choix de l'échantillon

La base de données ERIC (Educational Resources Information Center) a été sélectionnée parce qu'elle présente un bassin important de rapports d'évaluation de programme liés à l'éducation, dont plusieurs visant à prévenir la toxicomanie, la violence, la délinquance, le tabagisme, le stress ou la détresse émotionnelle et qui concernent principalement le domaine de l'éducation à la santé. Tout comme au premier volet de la recherche, la sélection des rapports s'est effectuée à partir de critères préalablement définis, à savoir :

- 1) exclusivement des rapports d'évaluation qui rapportent l'ensemble des étapes de la démarche (par exemple, la planification d'une évaluation ne pourrait être retenue);
- 2) s'inscrivant à l'intérieur de la période donnée (2000-2006).

Cinquante-sept (57) rapports d'évaluation ont été retenus selon ces critères comme échantillon initial. De ce nombre, dix-sept rapports ont été choisis de façon aléatoire pour des fins de préexpérimentation (10) et pour l'établissement d'un taux d'accord interjuge (7). Ces rapports n'ont pas été conservés pour l'analyse des données puisqu'ils ont plutôt servi aux chercheurs à se familiariser avec les grilles et qu'il était plus prudent de ne pas les comptabiliser, ce qui porte l'échantillon à quarante (40) rapports (dont le détail est présenté à l'appendice A).

3.2.3 La collecte des données

La cueillette de données s'appuie sur des grilles développées par l'équipe de recherche (*voir* appendice B). Les indicateurs de ces grilles sont présentés à cette section.

Les indicateurs des grilles d'analyse témoignant de l'application des composantes de la modélisation du processus spécifique à l'évaluation de programme sont présentés au tableau 3.1. Ils s'appuient sur les contributions théoriques de Scriven (1980), de Hurteau, (1991), de Fournier (1995), de Stake (2004) (*voir* section 2.3).

Tableau 3.1
Les indicateurs des grilles d'analyse témoignant de l'application de la modélisation

Composantes de la modélisation	Indicateurs : Sont mentionnés (exception : <i>Le respect de...</i>)
Description du programme	La clientèle du programme Ses objectifs Les caractéristiques de l'intervention Le cadre théorique sur lequel elle s'appuie
Problématique de l'évaluation	L'élément déclencheur de l'évaluation Le type de jugement attendu par le client Le type d'argumentation attendu par le client
Opérationnalisation de l'évaluation	Le/les question(s) et/ou l'/les objectif(s) de l'évaluation Le/les type(s) d'évaluation Le respect de la typologie de Rossi, Lipsey, Freeman (2004) Le respect du type de jugement attendu par le client
Choix stratégiques	Les critères Les standards La pondération des critères Le respect du type d'argumentation attendu par le client Les critères sont présentés au point de départ Les standards sont présentés au point de départ

Les indicateurs de la grille caractérisant les jugements et leurs fondements sont présentés au tableau 3.2. Ils sont inspirés des contributions théoriques d'Arens (2006) et de Toulmin (1993) (*voir* section 2.4).

Tableau 3.2
Les indicateurs de la grille d'analyse caractérisant le jugement et ses fondements

Caractéristiques du jugement	Indicateurs
Suggérés par les constats d'Arens (2006)	Les questions (ou objectifs si absence de questions) trouvent-elles leurs réponses par des déclarations?
	Les déclarations sont-elles synthétisées pour produire un jugement global?
	La façon dont cette synthèse est effectuée est-elle documentée?
	Des recommandations concernant la programmation sont-elles émises?
	Les recommandations concernent-elles l'aspect du programme à l'évaluation?
Suggérés par le modèle de Toulmin (1993)	Le jugement (ou déclarations si absence de jugement) est nuancé ?
	Les choix stratégiques sont-ils justifiés?
	Le rapport présente-t-il une justification concernant le choix des critères
	Le rapport présente-t-il une justification concernant le choix des standards?
	Sur quoi s'appuie cette justification?
	La justification des critères s'appuie sur...
	La justification des standards s'appuie sur ...

Du côté opérationnel, il y a eu obtention d'un taux d'accord interjuge moyen d'au moins 80% pour les rapports sélectionnés à cet effet. Des données qualitatives précisant la nature de l'information ont également été recueillies afin d'enrichir les fréquences présentées.

Pour explorer les éléments qui semblent faire en sorte que le jugement soit suffisamment fondé pour être accepté, la question suivante a été systématiquement répondue après la lecture de chacun des rapports d'évaluation : *Instinctivement, qu'est-ce qui fait en sorte que le lecteur adhère ou non au jugement?* Elle constitue le dernier indicateur de la grille d'analyse qui documente le jugement et ses fondements. Elle permet l'exploration d'éléments nouveaux qui pourraient influencer la crédibilité du jugement.

3.3 L'analyse des données

Puisque les analyses retenues favorisent l'atteinte des trois objectifs annoncés, elles seront présentées d'après ces trois objectifs.

Le premier objectif de recherche vise à *établir dans quelle mesure la modélisation du processus spécifique à l'évaluation de programme se retrouve au sein des pratiques évaluatives*. Il est atteint en présentant les fréquences obtenues à chacun des indicateurs témoignant de l'application des différentes composantes de la modélisation du processus spécifique à l'évaluation de programme (voir tableau 3.1). Il a été ainsi possible de déterminer dans quelle proportion les composantes sont utilisées pour structurer les démarches d'évaluation de programme analysées. De plus, des données qualitatives précisant la nature de l'information ont également été recueillies afin d'enrichir ces fréquences.

Le deuxième objectif vise à *caractériser les jugements qui y sont formulés*. Il est atteint en présentant les fréquences obtenues à chacun des indicateurs caractérisant les jugements et leurs fondements (voir tableau 3.2). Il a été ainsi possible de déterminer dans quelle proportion les différents indicateurs caractérisant le jugement sont présents à l'intérieur des démarches d'évaluation de programme analysées. De plus, pour explorer les éléments qui semblent faire en sorte que le jugement soit suffisamment fondé pour être accepté, les

impressions des chercheurs concernant ces éléments ont été systématiquement relus afin de déceler s'il s'en dégageait une tendance retrouvée dans la majorité des rapports.

Enfin, le troisième objectif vise à *étudier la contribution de la modélisation à la formulation d'un jugement fondé*. Il est réalisé au moyen d'une combinaison de méthodes quantitatives et de méthodes qualitatives. En ce qui concerne l'analyse quantitative, elle a permis d'analyser la relation existant entre les résultats des deux premiers objectifs, soit l'application de la modélisation (objectif 1) et la formulation d'un jugement fondé (objectif 2). Des tests statistiques selon la distribution du khi carré ont ainsi été effectués afin de percevoir la relation entre les fréquences témoignant de l'application de la modélisation (résultats objectif 1) et ceux qui caractérisent les jugements (résultat objectif 2). Chacune des fréquences aux indicateurs du tableau 3.1 (témoignant de la présence de l'application de la modélisation) ont alors été mises en relation avec celles aux indicateurs du tableau 3.2 (caractérisant le jugement et ses fondements). Quant à l'analyse qualitative, elle a plutôt permis l'examen de la contribution, et non de la relation, de la modélisation à la formulation d'un jugement fondé. Cet examen s'est effectué en examinant les résultats qualitatifs des jugements (indicateurs 11 et 12 du tableau 3.2) pour y déceler si des éléments de la modélisation y étaient repris.

3.4 Les forces de l'étude

Les choix méthodologiques qui ont façonné l'élaboration du plan de recherche présentent plusieurs forces dont les plus importantes seront présentées ici.

Premièrement, comme au premier volet de l'étude, c'est le contenu des rapports d'évaluation qui sera analysé plutôt que, par exemple, des entrevues auprès des praticiens. Ce choix méthodologique présente l'avantage indéniable d'analyser les pratiques en place et non pas la perception qu'en ont les praticiens.

Deuxièmement, contrairement au premier volet du programme de recherche qui a analysé des rapports d'évaluation ayant fait l'objet d'un article à l'intérieur d'une revue scientifique, de rapports d'évaluation originaux ont cette fois été à l'étude. Ces rapports risquent donc davantage d'être exempts de transformations importantes afin de se conformer aux exigences de la publication scientifique, permettant ainsi d'obtenir un portrait plus juste et plus exhaustif de la pratique.

Troisièmement, puisque l'échantillon des rapports constitue des évaluations de programme appartenant au domaine de l'éducation, les résultats générés ont davantage de chance d'être utilisés par ce domaine que si cette étude avait analysé des rapports d'évaluation liés à des domaines variés.

Quatrièmement, la validité de cette approche s'appuie sur les considérations de Rourke et Anderson (2004) qui énumèrent les étapes assurant une validité théorique au protocole d'analyse de contenu. La première étape, intitulée *identifier la raison pour laquelle les données seront recueillies* se retrouve vraisemblablement à l'intérieur de la problématique (voir chapitre 2) de cette recherche. La deuxième étape, intitulée *identifier les critères qui représenteront l'objet à l'étude* a également été réalisée lors de la construction des grilles d'analyse et s'appuie sur les contributions théoriques de Scriven (1980), de Hurteau, (1991), de Fournier (1995), de Stake (2004), d'Arens (2006) et de Toulmin (1993). La troisième étape, intitulée *étudier les catégories et les critères sélectionnés* s'est réalisée lors de la validation du contenu des grilles par la professeur Marthe Hurteau et le professeur Sylvain Houle, tous deux professeurs à l'Université du Québec à Montréal, dont l'expertise est reconnue dans le domaine de l'évaluation de programme. La quatrième étape, intitulée *mettre à l'essai la grille d'analyse*, s'est effectuée lors de la préexpérimentation des grilles sur les 10 rapports sélectionnés à cet effet qui a permis la création de la dernière version des grilles et par l'établissement d'un taux d'accord interjuge supérieur à 80% pour 7 autres rapports sélectionnés à cet effet. Enfin, la dernière étape, intitulée *développer des outils qui orientent l'administration, la cueillette et l'interprétation des données*, s'est vu amorcée lors du premier volet de l'étude et se continue par son approfondissement au présent volet du programme de recherche.

Finalement, Neuendorf (2002) dans De Wever (2006) précise que la validité interne de l'analyse de contenu se retrouvera principalement dans sa capacité à faire correspondre de manière systématique les éléments de la grille d'analyse et la théorie et que la validité externe sera, quant à elle, assurée par la possibilité de reproduire les données. Puisque ces préoccupations se retrouvent à l'intérieur des méthodes proposées (grilles explicites dont les indicateurs s'appuient sur des éléments théoriques), cette recherche présente ainsi plusieurs éléments qui traduisent un souci d'assurer une validité, autant interne qu'externe, à la démarche.

3.5 Les limites de l'étude

Les limites de cette étude se doivent également d'être présentées afin de nuancer de quelle manière les choix méthodologiques qui ont façonné l'élaboration du plan de recherche permettent de répondre à la question de recherche, soit *Dans quelle mesure l'application du processus spécifique à l'évaluation de programme contribue-t-elle à formuler un jugement fondé?*

Tout d'abord, si les rapports analysés à partir de ERIC (Educational Resources Information Center) présentent l'avantage d'être uniformes quant à la nature des contenus, il n'en demeure pas moins que ce choix présente un biais parce qu'il ne favorise pas une représentation exhaustive de la pratique.

Aussi, les rapports, qui constituent des comptes rendus écrits de l'évaluation ne présentent pas l'ensemble des choix relatifs à l'évaluation et la manière dont s'est déroulée pratiquement l'évaluation. Cette limite affecte particulièrement l'analyse des jugements et de leurs fondements qui se fait donc à partir de ce qui est écrit dans le rapport plutôt qu'en considérant la perception réelle des clients de l'évaluation qui ne pourrait être perçue qu'en effectuant des entrevues.

De plus, le contenu des grilles d'analyse aurait pu être validé par des personnes à l'extérieur du groupe de recherche. Il est cependant à noter qu'il y a relativement peu d'individus qui s'intéressent à cette dimension de l'évaluation de programme et que la contribution externe a donc été limitée à cet égard.

3.6 Synthèse

En conclusion, cette recherche se situe au confluent de la recherche descriptive et exploratoire parce qu'elle utilise autant la déduction permettant l'analyse détaillée des pratiques actuelles que l'induction, laissant émerger des réflexions sur les éléments qui semblent faire en sorte que le jugement soit suffisamment fondé pour être accepté. De plus, elle utilise aussi bien des méthodes quantitatives lorsqu'elle établit les fréquences selon lesquelles se retrouvent les différents indicateurs que des méthodes qualitatives lorsqu'elle précise la nature du contenu de certains indicateurs. Malgré leurs divers ports d'attache, ces choix méthodologiques s'harmonisent pour se consacrer à leur objectif commun, c'est-à-dire répondre à la question de recherche à partir de résultats les plus fidèles et valides possibles.

CHAPITRE IV

LES RÉSULTATS

Ce quatrième chapitre présente les résultats obtenus d'après le plan de recherche visant à répondre à la question suivante: *Dans quelle mesure l'application du processus spécifique à l'évaluation de programme contribue-t-elle à formuler un jugement fondé?*

Cette modélisation, illustrée à la figure 4.1, se divise en plusieurs composantes, qui sont elles-mêmes rendues opérationnelles au moyen de plusieurs indicateurs. Si chacune de ces composantes s'avère pertinente, parce que justifiée théoriquement à l'intérieur du cadre théorique de cette recherche, il n'en demeure pas moins que leur contribution à la formulation d'un jugement fondé n'a jamais été établie au moyen d'un protocole expérimental. C'est cet approfondissement qui constituera le cœur de la présente section et il peut se traduire par les trois objectifs suivants :

- 1- Établir dans quelle mesure la modélisation du processus spécifique à l'évaluation de programme se retrouve au sein des pratiques évaluatives;
- 2- Caractériser les jugements qui y sont formulés;
- 3- Étudier la contribution de la modélisation à la formulation d'un jugement fondé.

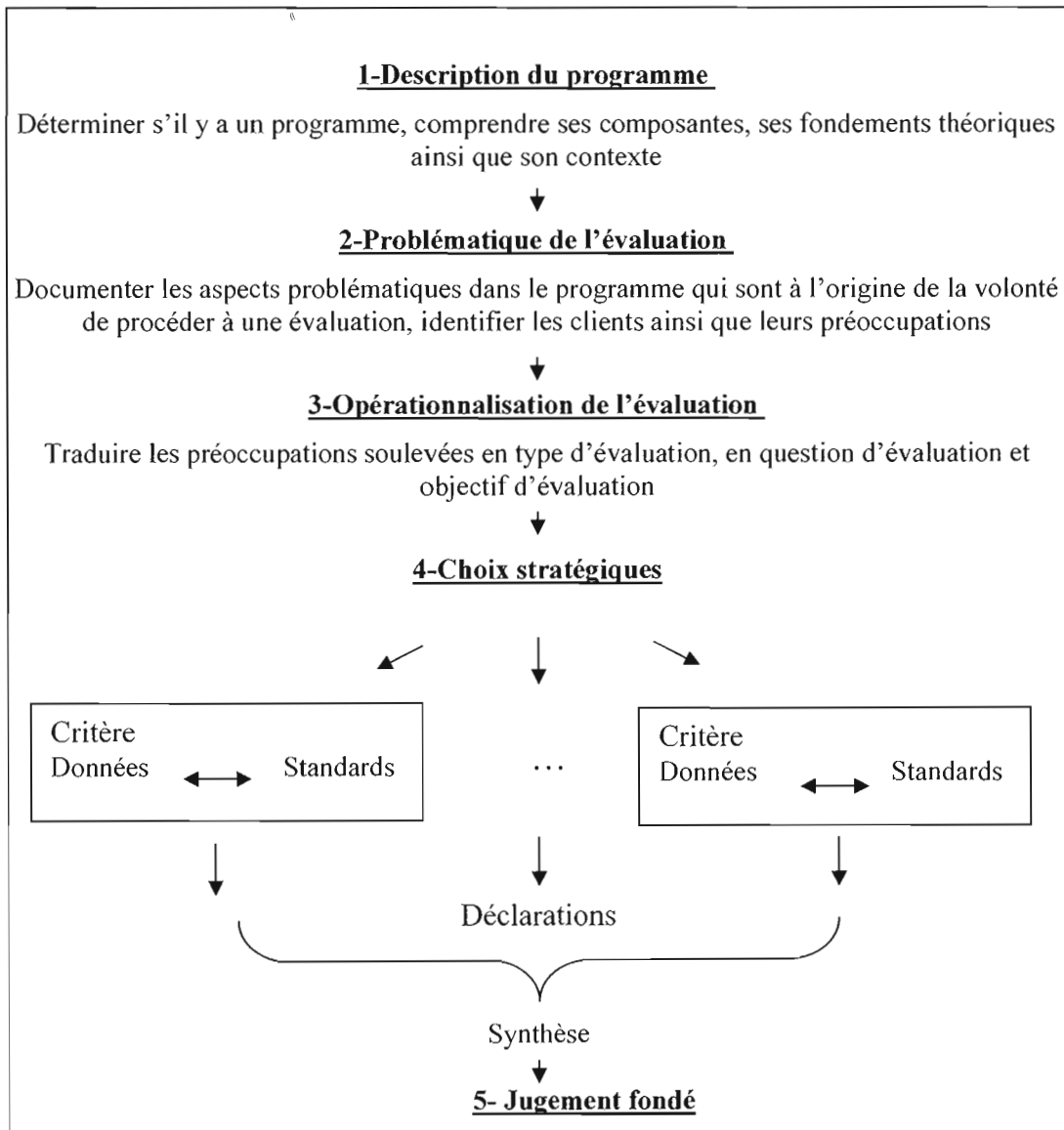


Figure 4.1 La modélisation du processus spécifique à l'évaluation de programme: éléments et interrelations. (Hurteau et Houle, 2006)

4.1 Objectif 1 : L'application de la modélisation

Les résultats au premier objectif de recherche, c'est-à-dire *établir dans quelle mesure la modélisation du processus spécifique à l'évaluation de programme se retrouve au sein des pratiques évaluatives* (voir figure 4.2) sont présentés à cette section.

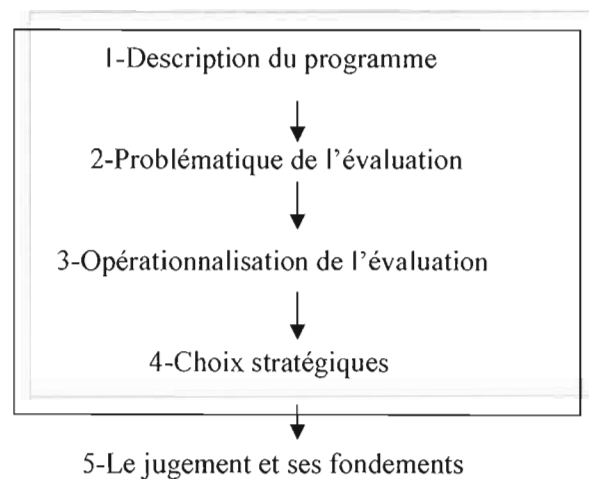


Figure 4.2 Les quatre premières composantes de la modélisation sont développées au premier objectif.

C'est ainsi que les fréquences obtenues aux différents indicateurs qui leur sont conceptuellement rattachés selon la modélisation de Hurteau, Houle (2006) (voir tableau 3.1) sont dévoilées à cette section. De plus, des données qualitatives précisant la nature de l'information ont également été recueillies afin d'enrichir les fréquences présentées.

4.1.1 La description du programme

Les résultats concernant la première composante de la modélisation, soit la *Description du programme*, référant aux aspects qu'il est nécessaire de documenter afin de s'assurer qu'il y a bien présence d'un programme (voir section 2.3.1), sont présentés au tableau 4.1.

Tableau 4.1
Analyse du contenu lié à la *Description du programme*

<i>Est-ce que le rapport décrit le programme à évaluer?</i>	Oui, explicitement	Non, il faut le déduire ou ce n'est pas mentionné
1) La clientèle est mentionnée	39 (97,5%)	1 (2,5%)
2) Les objectifs du programme sont mentionnés	31 (77,5%)	9 (22,5%)
3) Les caractéristiques de l'intervention sont mentionnées	37 (92,5%)	3 (7,5%)
4) Le cadre théorique sur lequel s'appuient les activités du programme est mentionné	15 (37,5%)	25 (62,5%)

Selon ces résultats, la clientèle à qui s'adresse le programme (98%) et les caractéristiques de l'intervention (93%) sont mentionnées à l'intérieur de la quasi-totalité des rapports. Les objectifs du programme (78%) le sont dans une proportion moindre et le cadre théorique sur lequel s'appuient les activités du programme (38%) l'est, enfin, dans une proportion encore plus faible.

Des données qualitatives viennent préciser la nature de ces résultats. En effet, non seulement la fréquence selon laquelle se retrouve un cadre théorique est faible (38%), mais il est très peu développé lorsque présent, se limitant majoritairement à la mention de lignes directrices orientant la philosophie du programme.

Il ressort ainsi de ces analyses que la *Description du programme* semble présente à l'intérieur des pratiques évaluatives étudiées bien que le cadre théorique sur lequel s'appuient les activités du programme ne soit pas un élément systématiquement documenté.

4.1.2 La problématique de l'évaluation

Les résultats concernant la deuxième composante de la modélisation, soit la *Problématique de l'évaluation* qui identifie globalement les besoins auxquels l'évaluation de programme devra répondre (voir section 2.3.2), sont présentés au tableau 4.2.

Tableau 4.2
Analyse du contenu lié à la *Problématique de l'évaluation*

<i>Est-ce que le rapport présente ce à quoi l'évaluation de programme devra répondre ?</i>	Oui, explicitement ou implicitement (il est possible de le déduire de façon évidente)	Non, ce n'est pas précisé
1) L'élément déclencheur de l'évaluation est précisé	17 (42,5%)	23 (57,5%)
2) Le rapport précise le type de jugement attendu par le client	6 (15%)	34 (85%)
3) Le rapport précise le type d'argumentation attendu par le client	3 (7,5%)	37 (92,5%)

Selon ces résultats, l'élément déclencheur de la démarche évaluative (ex. : politique d'évaluation, problème rencontré, besoin d'information, etc.), est précisé dans 43% des cas;

le type de jugement attendu par le client est rarement précisé (15%) et, finalement, le type d'argumentation attendu par le client est encore plus rarement précisé (8%).

Des données qualitatives viennent préciser la nature de ces résultats. En effet, l'élément déclencheur (43%) est très peu développé lorsque présent. Il se limite majoritairement à la mention d'une politique d'évaluation stipulant, par exemple, qu'une évaluation du programme doit être effectuée chaque année.

Il ressort ainsi de ces analyses que la *Problématique de l'évaluation* semble globalement peu présente à l'intérieur des rapports étudiés, pour ne pas dire inexistante, selon l'appui des données qualitatives.

4.1.3 L'opérationnalisation de l'évaluation

Les résultats concernant la troisième composante de la modélisation, soit l'*Opérationnalisation de l'évaluation* qui permet de traduire les besoins identifiés à la *Problématique de l'évaluation* en langage propre à l'évaluation de programme (voir section 2.3.3), sont présentés au tableau 4.3.

Tableau 4.3
Analyse du contenu lié à l'Opérationnalisation de l'évaluation

Est-ce que le rapport opérationnalise les éléments de la Problématique de l'évaluation?			
Est-ce que les indicateurs de l'Opérationnalisation de l'évaluation sont présents?	Oui, Explicitement	Non, il faut le déduire ou c'est absent	
1) Le/les type(s) d'évaluation effectué(s) est/sont spécifié(s)	4 (35%)	26 (65%)	
2) Le/les question(s) et/ou l'/les objectif(s) auquel /auxquels devra répondre l'évaluation est/sont spécifié(s)	25 (62,5%)	15 (37,5%)	
Est-ce que ces indicateurs sont liés selon la typologie de Rossi, Lipsey et Freeman (2004)?	Oui	Non	
		La typologie n'est pas respectée	Manque d'information (voir critère 1 et 2 de cette grille)
3) Le/les type(s) d'évaluation et les questions ou objectifs sont liés selon la typologie de Rossi, Lipsey et Freeman (2004)	7 (17,5%)	1 (2,5%)	32 (80%)
	8 (20%)		
Est-ce que ces indicateurs tiennent compte des éléments de la problématique?	Oui	Non	
		Ils n'en tiennent pas tous compte	Manque d'information (voir critère 2 de la grille B ou critères 1 et 2 de cette grille))
4) Le/les type(s) d'évaluation et le/les question(s) ou objectif(s) tiennent compte du type de jugement attendu par le client	2 (5%)	2 (5%)	36 (90%)
	4 (10%)		

Selon ces résultats, il n'y a qu'un tiers des rapports qui spécifient le type d'évaluation effectué (35%) bien qu'il soit possible, dans les autres cas, de le déduire à la lecture du rapport d'après les choix de l'évaluateur. Il n'en demeure pas moins que les deux tiers des rapports spécifient des questions ou des objectifs d'évaluation (63%). Aussi, sur les 8

rapports qui fournissent assez d'information en mentionnant à la fois le type d'évaluation et des questions/objectifs d'évaluation, 7 d'entre eux respectent la typologie de Rossi, Lipsey et Freeman (2004) qui les lie logiquement. Enfin, sur les 4 cas qui mentionnent le type de jugement attendu par le client à la *Problématique de l'évaluation*, la moitié des cas (2) en tiennent compte à l'intérieur de l'*Opérationnalisation de l'évaluation*.

Des données qualitatives viennent préciser la nature de ces informations. En effet, il est à noter que, selon les différents rapports, le sens attribué à un même type d'évaluation peut sensiblement varier. Ainsi, certains évaluateurs annoncent une évaluation d'impact en présentant des critères et un développement appartenant davantage à une évaluation de l'implantation. Il ne semble donc pas, tel que retrouvé à l'intérieur des rapports analysés, y avoir un consensus quant au sens exact attribué aux différents types d'évaluation. Ce même propos s'applique aux questions et objectifs d'évaluation dont le rôle varie selon l'utilisation des évaluateurs. Il ressort ainsi de ces analyses que les évaluateurs utilisent des éléments de l'*Opérationnalisation de l'évaluation* sans pour autant le faire de façon systématique.

4.1.4 Les choix stratégiques

Les résultats concernant la quatrième composante de la modélisation, soit les *Choix stratégiques* qui permettent d'établir les critères et les standards sur lesquels se basera le jugement (voir section 2.3.4), sont présentés au tableau 4.4.

Tableau 4.4
Analyse du contenu lié aux *Choix stratégiques de l'évaluation*

Est-ce que le rapport présente la tactique utilisée pour porter un jugement?				
Est-ce que les indicateurs des choix stratégiques sont présents?	Oui, explicitement ou implicitement (il est possible de le déduire de façon évidente)		Non	
	Entièrement	En partie	Non, avec justification	Non, sans justification
1) Les critères (éléments de l'objet qui seront à l'évaluation) sont présents	40 (100%)	0 (0%)	0 (0%)	0 (0%)
2) Les standards (points de coupure entre un jugement favorable et un jugement défavorable) sont présents	5 (12,5%)	18 (45%)	2 (5%)	15 (37,5%)
	23 (57,5%)			
3) Les critères ont une pondération relative qui leur est attribuée	0 (0%)	0 (0%)	4 (10%)	36 (90%)
Est-ce que ces indicateurs tiennent compte des éléments de la problématique?	Oui		Non	
			Ils n'en tiennent pas compte	Manque d'information (voir critère 3 de la grille B ou critères 1 et 2 de celle-ci)
4) Les critères et les standards tiennent compte du type d'argumentation attendu par le client	3 (7,5%)		0 (0%)	37 (92,5%)
5) Est-ce que les critères présents sont annoncés « au point de départ » ?				
Les critères sont présentés de façon à ce que tous les aspects de l'objet qui seront à l'évaluation soient présentés avant l'analyse des résultats.			23 (57,5%)	
Certains critères émergent en cours de route, les aspects de l'objet qui seront à l'évaluation sont alors modifiés avec l'arrivée de nouvelles informations.			12 (30%)	
Tous les critères émergent en cours de route, les aspects de l'objet qui seront à l'évaluation sont alors inclus avec l'arrivée de nouvelles informations.			5 (12,5%)	
Manque d'information - voir critère 1 de cette grille			0	
6) Est-ce que les standards présents sont annoncés « au point de départ » ?				
Les standards sont présentés de façon à ce que tous les points de coupure qui détermineront le niveau de performance à atteindre soient présentés implicitement ou explicitement avant l'analyse des résultats.			17 (42,5%)	
Certains standards émergent en cours de route, ces points de coupures sont alors précisés avec l'arrivée de nouvelles informations.			4 (10%)	
Tous les standards émergent en cours de route, les points de coupures sont alors précisés avec l'arrivée de nouvelles informations.			2 (5%)	
Manque d'information - voir critère 2 de cette grille			17 (42,5%)	

Selon ces résultats, la totalité des rapports d'évaluation a utilisé des critères d'évaluation bien qu'il n'y ait que la moitié des rapports qui les présentent entièrement avant l'analyse des résultats (58%). Les autres les présentent partiellement avant l'analyse des résultats (30%) ou encore n'en présentent aucun avant l'analyse des résultats (13%). En ce qui concerne l'utilisation de standards, seulement 2 rapports sur les 17 n'ayant pas utilisé de standards, justifient ce fait. Ils évoquent diverses considérations, dont principalement l'absence de standards préétablis. Sur les 23 rapports qui en ont utilisé, 5 en ont systématiquement utilisé et 18, partiellement, c'est-à-dire que certains critères n'étaient pas liés à un standard. Il est également possible de remarquer qu'ils sont généralement présentés plus tôt que les critères, car de ces 23 rapports, la plupart (17) les présentent entièrement avant l'analyse des résultats.

Concernant le lien entre cette composante et les éléments de la *Problématique de l'évaluation*, sur les 3 cas qui précisent le type d'argumentation attendu par le client de l'évaluation à leur *Problématique de l'évaluation*, la totalité (3) en tient compte à l'intérieur de leurs *Choix stratégiques*.

Enfin, l'attribution d'une pondération aux critères pour produire un jugement global est inexistante (0%) même si ce fait est parfois justifié implicitement (10%). En effet, certains rapports présentent une méthodologie utilisant des schémas expérimentaux qui se basent sur un seul critère général pour émettre leur jugement et ne tireraient ainsi aucun avantage à en utiliser.

Il ressort de ces analyses que les *Choix stratégiques* semblent se retrouver à l'intérieur des pratiques évaluatives étudiées, bien que certains indicateurs ne soient pas utilisés de manière systématique, dont les standards, la pondération des critères et l'annonce de ces deux éléments au point de départ.

4.1.5 Conclusion (Objectif 1)

Il ressort de cette analyse que les quatre premières composantes de la modélisation du processus spécifique à l'évaluation de programme ne se retrouvent pas dans une proportion homogène à l'intérieur des rapports étudiés. À cet égard, le tableau 4.5 effectue une synthèse des résultats présentés en affichant les moyennes des fréquences obtenues pour chacune des composantes de la modélisation. Ces moyennes ont été obtenues en effectuant la somme des fréquences de tous les indicateurs d'une composante divisée ensuite par le nombre d'indicateurs que possède cette composante.

Tableau 4.5
Fréquence moyenne des indicateurs des composantes liées au premier objectif de recherche

Composante de la modélisation	-1- Description du programme	-2- Problématique de l'évaluation	-3- Opération- nalisation de l'évaluation	-4- Choix stratégiques
Fréquence moyenne (%)	76,25%	21,66%	48,75%	59,5%

Il est à noter que ces moyennes doivent être interprétées avec précaution, car il a fallu exclure certains indicateurs pour lesquels certains rapports ne présentaient pas assez d'information pour juger de leur présence (existence d'une case « Manque d'information » à l'indicateur). L'inclusion des fréquences leur étant rattachées aurait faussé la moyenne en n'incluant pas la totalité des rapports.

Ces résultats mettent toutefois en lumière que la plupart des rapports ne documentent pas la *Problématique de l'évaluation* (22%). Ils montrent également que l'*Opérationnalisation de l'évaluation*, dont l'analyse qualitative révèle des irrégularités au niveau de son utilisation, n'est pas réalisée de façon systématique (49%). La *Description du programme*, est relativement fréquente (76%), suivie des *Choix stratégiques* (60%).

4.2 Objectif 2 : Les caractéristiques des jugements et de leurs fondements

Les résultats au deuxième objectif de recherche, c'est-à-dire *caractériser les jugements qui y sont formulés*, (voir figure 4.3), sont présentés à cette section. Les résultats concernant le jugement fondé revêtent une importance particulière car ils constituent la raison d'être de toute évaluation de programme.

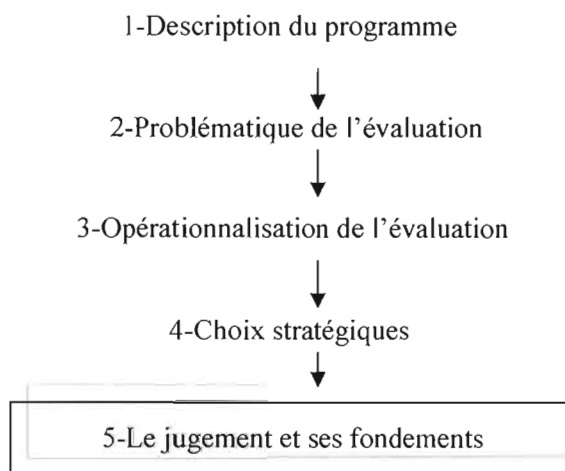


Figure 4.3 Les jugements, la finalité de la modélisation, sont caractérisés au deuxième objectif.

C'est ainsi que les fréquences obtenues aux différents indicateurs qui lui sont conceptuellement rattachés sont dévoilées selon l'apport respectif d'Arens (2006) et de Toulmin (1993). Ces résultats seront également accompagnés des impressions des chercheurs sur les éléments qui semblent faire en sorte que le jugement soit suffisamment fondé pour être accepté.

4.2.1 Le jugement et ses fondements; d'après les constats de Arens (2006)

Un rappel des constats d'Arens (2006) qui ont inspiré certains indicateurs caractérisant le jugement fondé est présenté à la figure 4.4.

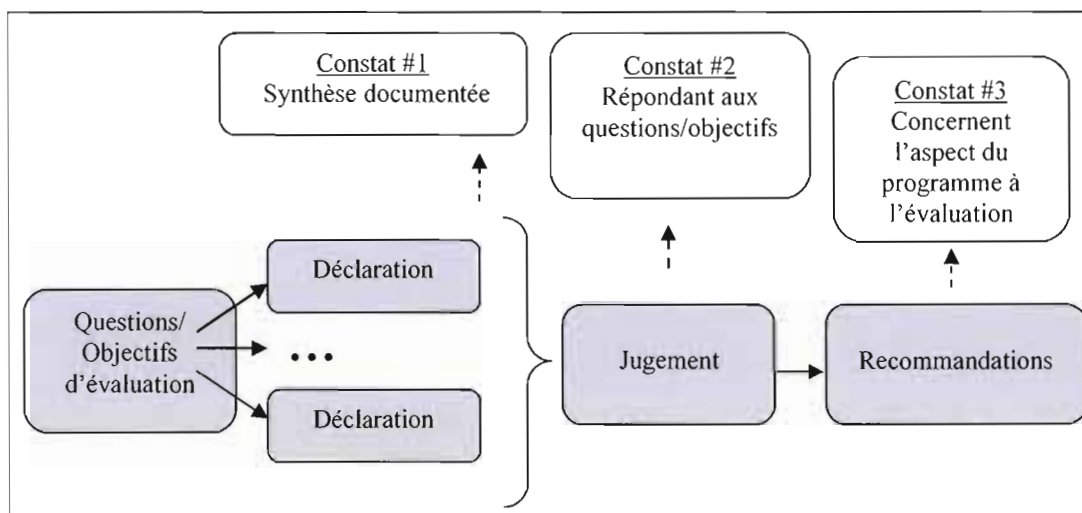


Figure 4.4 Les indicateurs témoignant d'un jugement fondé (en blanc) inspirés des constats d'Arens (2006) incorporés à sa représentation de la construction d'un jugement (en gris).

Les fréquences obtenues à chacun des indicateurs leur étant conceptuellement rattachés (voir tableau 3.2) sont présentées au tableau 4.6.

Tableau 4.6
Analyse du contenu lié au *jugement et ses fondements* d'après l'apport d'Arens (2006)

<i>D'après l'apport d'Arens (2006), comment se présentent le jugement et ses fondements?</i>			
<i>Les questions trouvent-elles une réponse?</i>	Oui	Non	Manque d'information
1) Les questions (ou objectifs si absence de questions) trouvent-elles leurs réponses?	19 (47,5%)	6 (15%)	15 (37,5%)
	25 (62,5%)		
<i>Est-ce qu'un jugement est émis et comment?</i>	Oui	Non	Manque d'information
2) Les déclarations sont-elles synthétisées pour produire un jugement global?	20 (50%)	20 (50%)	0 (0%)
3) La façon dont cette synthèse est effectuée est-elle documentée?	0 (0%)	20 (50%)	20 (50%)
<i>Les recommandations sont-elles appuyées?</i>	Oui	Non	Manque d'information
4) Des recommandations concernant la programmation sont-elles émises?	18 (45%)	22 (55%)	0 (0%)
5) Les recommandations concernent-elles l'aspect du programme à l'évaluation?	5 (12,5%)	13 (32,5%)	22 (55%)
	18 (45%)		

Selon ces résultats, sur les 25 rapports qui mentionnent une question (ou objectifs si absence de questions), 19 y répondent par des déclarations. Bien que ce nombre soit relativement élevé, le fait que les autres rapports (6) émettent des questions (ou objectifs si absence de questions) qui ne trouvent pas de réponse à l'intérieur du rapport doit cependant être mis en relief. De plus, la moitié des rapports (50%) synthétisent les différentes déclarations en un jugement global concernant le programme à l'évaluation. En se basant sur le fait que le jugement constitue la raison d'être d'une démarche d'évaluation de programme, ce résultat est inquiétant et indique, à tout le moins, que les évaluateurs ne transfèrent pas ce principe à leur pratique. De plus, aucun (0%) de ces rapports ne documente la façon dont la synthèse a été effectuée laissant au lecteur le soin d'extrapoler à ce sujet, confirmant les constats d'Arens (2006) à ce sujet (*voir la section 2.2.5*).

Ajoutons à cette analyse, celle des recommandations qui traduisent les jugements lorsque ces derniers ne sont pas explicitement présentés à l'intérieur du rapport. Les résultats démontrent que près de la moitié des rapports (45%) émettent des recommandations. Toutefois, sur les 18 rapports qui en émettent, 13 d'entre eux émettent des recommandations concernant un autre aspect du programme à l'évaluation.

4.2.2 Le jugement et ses fondements; d'après le modèle de Toulmin (1993)

Un rappel des considérations de Toulmin (1993) qui ont inspiré certains indicateurs caractérisant le jugement fondé est présenté à la figure 4.5.

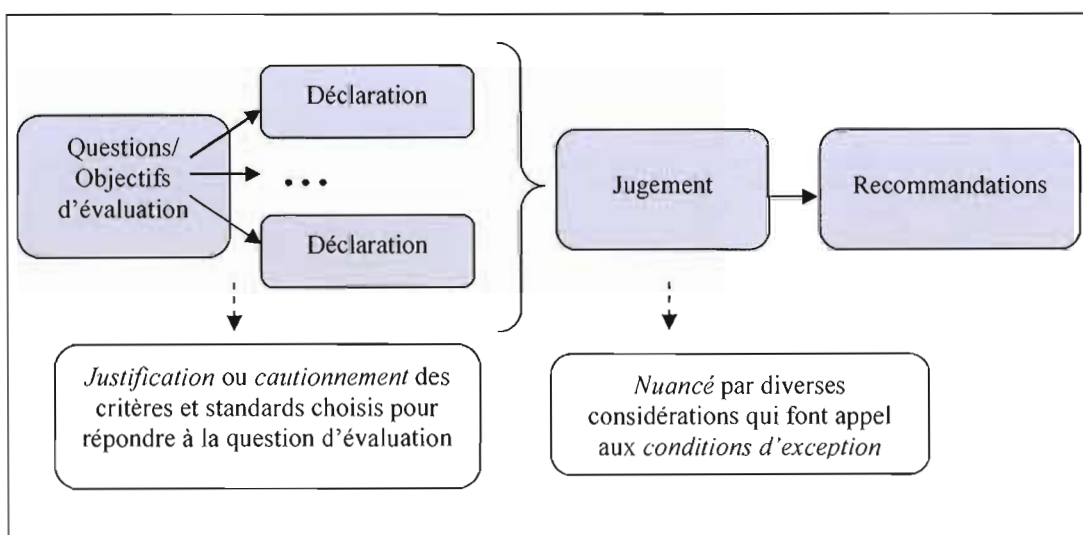


Figure 4.5 Les indicateurs témoignant d'un jugement fondé inspirés de la contribution de Toulmin (1993) (en blanc) incorporés à la représentation de la construction d'un jugement d'Arens (2006) (en gris).

Les fréquences obtenues à chacun des indicateurs leur étant conceptuellement rattachés (voir tableau 3.2) sont présentées au tableau 4.7. Elles sont accompagnées des impressions des chercheurs sur ce qui semble faire en sorte que le jugement soit suffisamment fondé pour être accepté.

Tableau 4.7Analyse du contenu lié au *jugement et ses fondements* d'après l'apport de Toulmin (1993)

D'après l'apport de Toulmin (1993), comment se présentent le jugement et ses fondements?				
Le jugement est-il nuancé?	Oui		Non	
6) Le jugement (ou déclarations si absence de jugement) est nuancé	15 (37,5%)		25 (62,5%)	
Les choix stratégiques sont-ils justifiés?	Oui (implicite ou explicite)		Non, ces éléments ne sont pas justifiés	
	Implicite	Explicite	Ils ne sont pas justifiés	Manque d'information
7) Le rapport présente-t-il une justification concernant le choix des critères?	29 (72,5%)	8 (20%)	3 (7,5%)	0 (0%)
	37 (92,5%)			
8) Le rapport présente-t-il une justification concernant le choix des standards?	18 (45%)	5 (12,5%)	0 (0%)	17 (42,5%)
	23 (57,5%)			
Sur quoi s'appuie cette justification?	Éléments (Il est possible d'avoir plusieurs sources de justification)			
9) La justification des critères s'appuie sur...	-Non justifiés (2) -La clientèle (14), les objectifs du programme (31), les caractéristiques de l'intervention (24), le contexte du programme (11), le cadre théorique du programme (3) -Résultats d'une évaluation précédente (1)			
10) La justification des standards s'appuie sur ...	-Absence de standard-Manque d'information (17) - Jugement de l'évaluateur (1) -Comparaison avec un groupe contrôle sans test statistique (5), comparaison avec un groupe contrôle avec tests statistiques (7), comparaison pré/post sans test statistique (2), comparaison pré/post avec test statistique (7), corrélation entre deux variables (2), -Données concernant programme semblable (3), moyenne nationale (3)			
Impressions sur ce qui semble faire en sorte que le lecteur adhère au jugement ou non				
11) Instinctivement, qu'est-ce qui semble faire en sorte que le lecteur adhère ou non au jugement?	- Ce serait la « force » du lien logique unissant les critères choisis à l'aspect du programme à l'évaluation, en d'autres mots, la bonne représentativité des critères selon le contexte du programme et celui de son évaluation qui semble faire en sorte que le lecteur adhère ou non au jugement. - La justification du choix des standards par principalement le recours à des considérations méthodologiques ferait également en sorte que le lecteur adhère ou non au jugement mais d'une façon beaucoup plus subtile, servant plutôt à le nuancer.			

D'après ces résultats, une proportion relativement faible des rapports nuance le jugement (ou déclaration si absence de jugement) par diverses considérations (38%) en évoquant principalement les forces et les limites méthodologiques de l'évaluation. Aussi, si dans la quasi-totalité des cas (37) le choix des critères est justifié, il n'en demeure pas moins qu'il l'est majoritairement de façon implicite (28), c'est-à-dire que les raisons pour lesquelles les critères ont été choisis peuvent être déduites à la lecture du rapport sans toutefois être présentées. Quant au choix des standards, le même phénomène est observé, c'est-à-dire qu'il est majoritairement justifié de façon implicite (18 rapports sur 23). Ce constat illustre le fait que c'est le lecteur qui doit déduire les raisons pour lesquelles l'évaluateur a choisi les critères et les standards. Conséquemment, ces raisons pourraient aussi varier selon différents lecteurs.

En ce qui concerne ce sur quoi s'appuie la justification implicite, c'est sur les éléments de la *Description du programme* à évaluer (la clientèle, les objectifs du programme, les caractéristiques de l'intervention, le contexte du programme et le cadre théorique du programme) que la majorité des évaluations appuie leurs critères. Un seul rapport déroge à cette tendance en justifiant ses critères par les résultats d'une évaluation précédente et les 3 autres ne présentent pas de justification à leurs critères, même en utilisant la déduction. Concernant le choix des standards, il est toujours justifié implicitement et s'appuie principalement sur des considérations méthodologiques issues de schémas expérimentaux. D'autres types de justifications sont également présents dont une référence aux données concernant un programme semblable (3 cas) et aux moyennes nationales (3 cas). Enfin, un rapport se détache de cette tendance et justifie ses standards en faisant appel à la qualité du jugement de l'évaluateur.

En résumé, peu de rapports mentionnent les raisons pour lesquelles l'évaluateur a choisi les critères et les standards (9 rapports pour les critères et 5 rapports pour les standards). Cependant, il est, hormis 3 cas, toujours possible de déduire ces raisons d'après les informations fournies à l'intérieur du rapport. Lorsqu'elles doivent être déduites, ces raisons illustrent que le choix des critères s'appuie principalement sur la description du programme à évaluer et le choix des standards s'appuie sur les schémas expérimentaux.

Dans un tout autre ordre d'idée, nous avons tenté d'établir si l'analyse des rapports permet de dégager une « impression » sur les éléments qui semblent faire en sorte que le jugement soit suffisamment fondé pour être accepté. Cette impression a été documentée de façon systématique en fournissant après la lecture de chacun des rapports une réponse à la question suivante: « Instinctivement, qu'est-ce qui semble faire en sorte que le lecteur adhère ou non au jugement? ». Même si nous reconnaissons que cette impression demeure très intuitive, il n'en demeure pas moins qu'elle peut s'avérer importante en débouchant sur de nouvelles pistes de recherche. Ainsi, il se dégage de toutes les impressions documentées que *c'est la « force » du lien logique unissant les critères choisis à l'aspect du programme à l'évaluation, en d'autres mots, la bonne représentativité des critères selon le contexte du programme et celui de son évaluation qui semble faire en sorte que le lecteur adhère ou non au jugement. Aussi, la justification du choix des standards par, principalement, le recours à des considérations méthodologiques, ferait également en sorte que le lecteur adhère ou non au jugement mais d'une façon beaucoup plus subtile, servant plutôt à le nuancer.*

4.2.3 Conclusion (Objectif 2)

Il ressort de cette analyse que le jugement et ses fondements, ne se présentent pas de manière uniforme. La figure 4.6 effectue une synthèse de ces résultats en présentant la distribution des rapports selon leurs résultats aux différents indicateurs. Elle reprend certains résultats déjà présentés mais cette fois-ci, à l'intérieur d'un diagramme en arbre. Le premier embranchement dévoile la distribution des résultats au deuxième indicateur du tableau 4.7, soit *si les déclarations sont synthétisées pour produire un jugement global sur l'objet*. Le deuxième embranchement présente la distribution des résultats au quatrième indicateur de ce même tableau, soit *si le jugement ou les déclarations évaluatives sont nuancés par diverses considérations*. Enfin, le troisième embranchement représente la distribution au premier indicateur de ce tableau, soit *si les questions et/ou objectifs trouvent leurs réponses par des déclarations*.

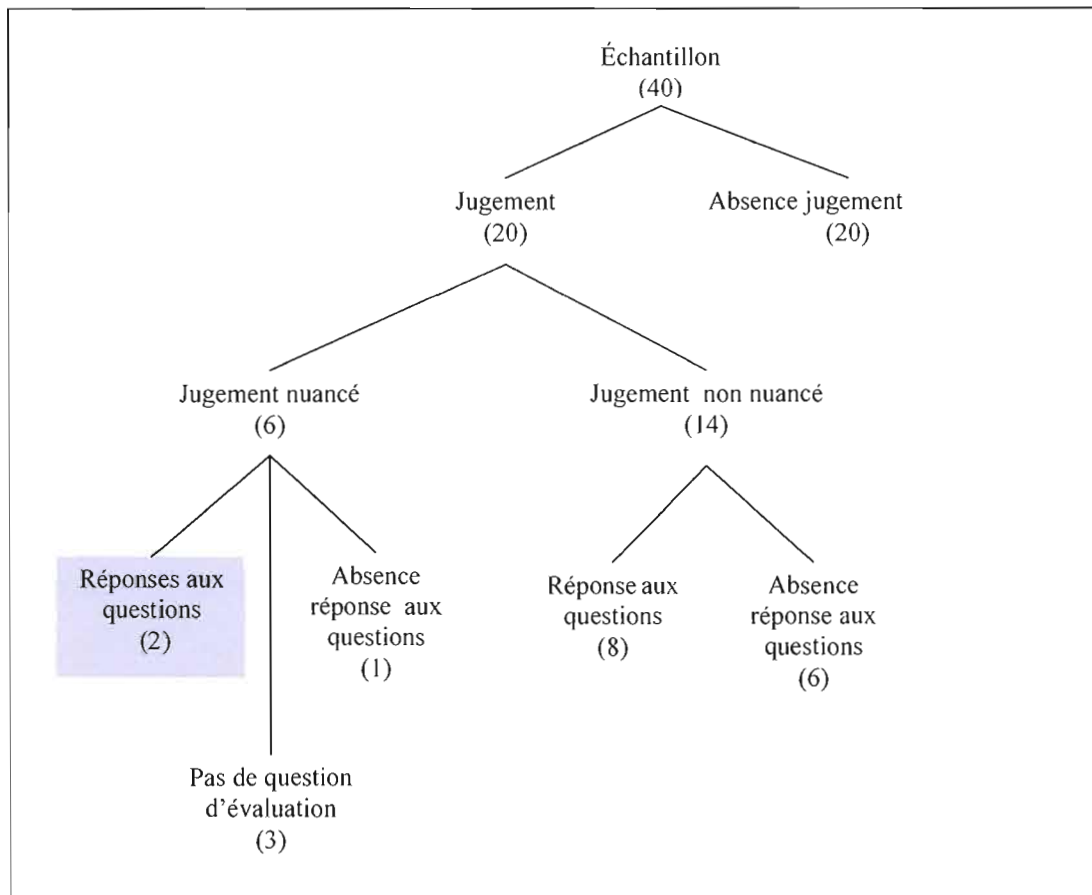


Figure 4.6 La distribution des cas selon les caractéristiques des jugements.

Ce diagramme met également en évidence qu'il y a seulement une infime portion des rapports (2 cas) qui produisent un jugement nuancé concernant l'aspect du programme à l'évaluation tout en répondant aux questions et/ou objectifs d'évaluation. Ce constat s'ajoute au fait qu'aucun rapport n'explique la manière dont les déclarations ont été synthétisées pour construire le jugement et que plusieurs rapports émettent des recommandations qui dépassent la portée du type d'évaluation effectué (13 rapports sur les 18 ayant émis des recommandations).

Enfin, ces résultats, qui caractérisent le jugement et ses fondements, permettront leur mise en relation avec ceux relatifs à l'application du processus spécifique à l'évaluation de

programme afin de répondre au troisième objectif de recherche, présenté à la section suivante.

4.3 Objectif 3 : Étudier la contribution de l'application de la modélisation à la formulation d'un jugement fondé

Les résultats au troisième objectif de recherche, c'est-à-dire *étudier la contribution de l'application de la modélisation à la formulation d'un jugement fondé* (voir figure 4.7) sont présentés à cette section.

Si les deux premiers objectifs étaient de nature purement descriptive, il n'en est pas de même pour le troisième objectif qui tente de cerner la contribution de l'application de la modélisation (résultats du premier objectif) à la formulation d'un jugement fondé (résultats de l'objectif précédent). Bien que différents auteurs appuient la présence de cette relation (Guba, 1972 et Scriven, 1995), aucune étude n'a encore tenté de la confirmer au moyen de données scientifiques. Pour percevoir cette relation, une analyse quantitative complétée d'un examen qualitatif a été effectuée.

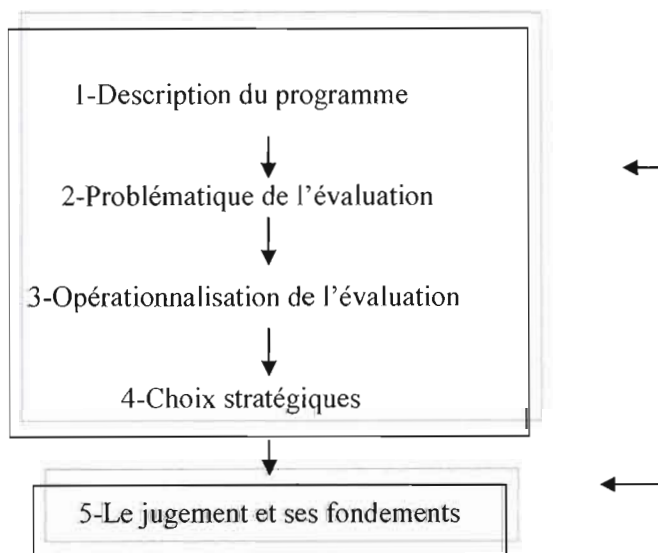


Figure 4.7 La relation à étudier au troisième objectif.

4.3.1 La mise en relation d'après les analyses quantitatives

En ce qui concerne l'analyse quantitative, des tests statistiques selon la distribution du khi carré ont été effectués afin de percevoir la relation entre les indicateurs témoignant de l'application de la modélisation du processus spécifique à l'évaluation de programme et ceux caractérisant les jugements. Chacun des indicateurs témoignant de la présence de l'application de la modélisation ont été mis en relation avec ceux caractérisant le jugement et ses fondements. Les analyses quantitatives selon la distribution du khi carré révèlent qu'il n'y a pas de relation statistiquement significative ($\alpha < 0,05$) entre les indicateurs témoignant de l'application de la modélisation et ceux caractérisant les jugements et leurs fondements.

4.3.2 L'analyse de la contribution selon un examen qualitatif

Selon un autre angle d'analyse, nous avons examiné les résultats qualitatifs des jugements (*voir* tableau 4.7) pour y déceler si des éléments de la modélisation y étaient repris. Cet examen indique que le contenu des indicateurs 9 et 11 reprend effectivement différentes composantes de la modélisation. Le tableau 4.8 les présente une seconde fois et les lie à la composante à laquelle ils correspondent. Il témoigne ainsi de la reprise de certaines composantes de la modélisation à l'intérieur des résultats qualitatifs caractérisant les jugements.

Tableau 4.8
Examen des résultats qualitatifs des jugements pour y déceler si des éléments de la modélisation y sont repris

Résultats qualitatifs liés au jugement fondé (extraits du tableau 4.7)	Composantes de la modélisation correspondantes
<p>Selon les impressions intuitives des chercheurs, qu'est-ce qui ferait en sorte que le lecteur adhère ou non au jugement?</p> <p>- Ce serait la « force » du lien logique unissant les critères choisis à l'aspect du programme à l'évaluation, ou en d'autres mots, la bonne représentativité des <u>critères(iii)</u> selon le <u>contexte du programme (i)</u> et <u>celui de son évaluation (ii)</u> qui semble faire en sorte que le lecteur adhère ou non au jugement.</p> <p>- La justification du <u>choix des standards</u> par principalement le recours à des considérations méthodologiques ferait également en sorte que le lecteur adhère ou non au jugement mais d'une façon beaucoup plus subtile, servant plutôt à le nuancer</p>	<p>i- Élément de la <i>Description du programme</i>,</p> <p>ii- Élément de la <i>Problématique de l'évaluation</i></p> <p>iii- Élément des <i>Choix stratégiques</i></p>
<p>La justification des critères s'appuie sur :</p> <p>-La clientèle (14 cas) ; -Les objectifs du programme (31 cas); -Les caractéristiques de l'intervention (24 cas); -Le contexte du programme (11 cas) ; -Le cadre théorique du programme (3 cas).</p>	<p>iv-Élément de la <i>Description du programme</i></p>

Par exemple, les résultats concernant ce qui ferait en sorte que le lecteur adhère ou non au jugement reprennent les éléments de trois composantes, soit les *Choix stratégiques*, la *Description du programme* et la *Problématique de l'évaluation*, en impliquant le choix des critères, le contexte du programme et le choix des standards (voir tableau 4.8 (i, ii et iii)). De plus, les indicateurs de la composante *Description du programme*, servent tous à justifier le choix des critères, un indicateur qualitatif caractérisant le jugement (voir tableau 4.8 (iv)). L'*Opérationnalisation de l'évaluation* semble toutefois faire exception, car elle est la seule composante qui ne semble pas posséder des éléments qui sont repris à l'intérieur de ces résultats. En examinant la nature de cette contribution, certains éléments de la modélisation peuvent être repris pour justifier les critères et pour faire en sorte que le lecteur adhère au

jugement. Il est ainsi possible de les réinvestir à l'intérieur d'une argumentation pour convaincre ou dissuader du bien-fondé du jugement.

4.3.3 Conclusion (Objectif 3)

En résumé, l'analyse quantitative ne révèle pas de relation entre l'application du processus spécifique à l'évaluation de programme et la formulation d'un jugement fondé. Cependant, l'examen qualitatif révèle que des éléments de la modélisation pourraient être réinvestis à l'intérieur d'une argumentation au jugement pour convaincre ou dissuader du bien-fondé du jugement. Les éléments de l'*Opérationnalisation de l'évaluation* semblent toutefois faire exception, car ils appartiennent à la seule composante qui n'apparaît pas avoir joué ce rôle.

CHAPITRE V

LA DISCUSSION

Les résultats ayant soulevé plusieurs questions concernant tant leur portée que leur pertinence, celles-ci seront discutées au présent chapitre. Elles sont formulées de la façon suivante :

-Quels sont les résultats les plus révélateurs de la pratique actuelle en évaluation de programme (objectif 1)?

-Dans quelle mesure les résultats concernant les caractéristiques des jugements et de leurs fondements (objectif 2) précisent-ils le malaise énoncé par Guba (1972) et Scriven (1995)?

-Quel est l'impact des résultats concernant la contribution de la modélisation à la formulation d'un jugement fondé (objectif 3) dans le domaine de l'évaluation de programme?

5.1 Quels sont les résultats les plus révélateurs de la pratique actuelle en évaluation de programme (objectif 1)?

Certains résultats méritent d'être soulignés ici comme étant des éléments qui dévoilent des aspects importants de la pratique en évaluation de programme. Ils pourraient ainsi constituer des pistes de réflexion pour la pratique.

1) 25 rapports sur 40 (63%) ne mentionnent pas de cadre théorique appuyant leur intervention (voir section 4.1.1)

Le fait que la majorité des rapports n'associe aucun cadre théorique à leur intervention pourrait être nuancé par une précision. Il est possible que certains évaluateurs aient pris connaissance du cadre théorique du programme, sans toutefois en avoir fait mention à l'intérieur de leur rapport. D'un autre côté, ce résultat pourrait illustrer un malaise relatif aux fondements théoriques des interventions des programmes sociaux actuellement implantés. Enfin, en considérant que cette composante pourrait être réinvestie à l'intérieur d'une argumentation pour convaincre ou dissuader du bien-fondé du jugement (voir section 4.3.2), il serait souhaitable d'approfondir les raisons de la faible fréquence de cet indicateur.

2) La problématique de l'évaluation, est la composante la moins présente (fréquence moyenne des indicateurs de 22%) (voir section 4.1.2).

La faible représentation de la *Problématique de l'évaluation*, pourrait, encore une fois, être expliquée par le fait que les évaluateurs aient pu connaître les attentes des clients (concernant le type de jugement et d'argumentation attendu) sans toutefois les inclure à l'intérieur de leurs rapports. Cependant, le fait qu'il n'y ait que 17 rapports (42,5%) qui précisent ce qui a déclenché l'évaluation en mentionnant principalement une politique d'évaluation (voir section 4.1.2) laisse sous-entendre que ce n'est pas une composante de la modélisation que les évaluateurs examinent en profondeur. Ce constat est d'autant plus préoccupant que les rapports étudiés s'adressent bel et bien aux clients de l'évaluation. Enfin, en considérant que cette composante pourrait être réinvestie à l'intérieur d'une argumentation pour convaincre ou dissuader du bien-fondé du jugement (voir la section 4.3.2), il serait également souhaitable d'approfondir les raisons de la faible fréquence des indicateurs de la *Problématique de l'évaluation*.

3) *L'Opérationnalisation de l'évaluation (Questions/Objectifs et type d'évaluation, dont l'analyse qualitative révèle des irrégularités au niveau de son utilisation, n'est pas effectuée de façon systématique (49%). (voir section 4.1.3)*

La confusion entourant les termes employés par les évaluateurs pour traduire les besoins des clients en langage d'évaluation témoigne que le langage propre à l'évaluation de programme n'est pas encore assez circonscrit pour être utilisé de façon systématique et homogène. Ce résultat appuie également les intuitions de certains auteurs qui croient que le développement de ce domaine se fera par l'enrichissement de son corpus théorique (Dubois et Marceau, 2005). Ainsi, ce domaine aurait tout avantage à établir un consensus quant au sens exact des différents types d'évaluation pour assurer une certaine cohérence à l'intérieur des pratiques. Ce même propos s'applique aux questions et objectifs d'évaluation dont le rôle varie selon l'utilisation des évaluateurs. Enfin, ce n'est qu'une fois la théorie derrière l'*Opérationnalisation de l'évaluation* et ses indicateurs, soit le type d'évaluation et les questions/objectifs d'évaluation, précisée qu'il sera possible de cerner sa contribution au sein du processus spécifique à l'évaluation de programme et, ultimement, à la formulation d'un jugement fondé.

4) *L'utilisation des standards n'est pas systématique (57,5%) (voir section 4.1.4)*

Même si les standards sont utilisés et que ceux qui n'en utilisent pas le justifient par l'absence de standards préétablis, ce résultat remet en cause le sens du mot « évaluer » à l'intérieur de l'expression « *évaluation* de programme » tel que décrit par Scriven (1980). En effet, selon les propos de cet auteur, la mise en relation des données avec des standards (point de coupure entre un jugement favorable et défavorable) serait indissociable de l'acte d'évaluer. Selon cette perspective, il serait intéressant de documenter davantage ce lien pour arriver à confirmer l'intuition que plusieurs chercheurs partagent à ce sujet, à savoir qu'il est nécessaire d'utiliser des standards afin de dépasser le niveau de la description pour produire un jugement, ultime finalité de l'évaluation de programme.

5) *Aucun rapport n'établit de « pondération » pour exprimer le poids relatif des différents critères utilisés (voir section 4.1.4)*

Ce résultat appuie le constat d'Arens (2006) qui stipule que la pondération des critères, indicateurs des *Choix Stratégiques*, n'est nullement utilisée en pratique. Pourquoi est-ce ainsi malgré le fait que la contribution de cet élément au processus spécifique à l'évaluation de programme semble claire pour les chercheurs qui étudient la démarche de construction d'un jugement dont Scriven (1995) et Arens (2006)? Seul un examen plus approfondi du processus qui consiste à synthétiser les différentes déclarations pour produire un jugement pourrait le révéler.

Enfin, ces résultats témoignent d'un manque de consensus quant à l'utilisation des éléments qui sont spécifiques à l'évaluation de programme à l'intérieur des rapports d'évaluation. Il est cependant nécessaire de nuancer ces propos en précisant que cette étude a eu pour objet les pratiques évaluatives telles qu'écrites à l'intérieur de rapports d'évaluation (40). Conséquemment, l'absence de certaines composantes ne veut pas dire qu'elles n'ont pas été prises en considération, mais bien qu'elles n'ont pas été rapportées. Sont-elles réellement absentes ou ont-elles été considérées comme moins nécessaires au rapport? Seules des entrevues avec les évaluateurs pourraient fournir des réponses à ces questions. Ces entrevues pourraient également préciser la contribution des clients de l'évaluation aux différentes composantes du processus spécifique à l'évaluation de programme.

5.2 Dans quelle mesure les résultats concernant les caractéristiques des jugements et de leurs fondements (objectif 2) précisent-ils le malaise énoncé par Guba (1972) et Scriven (1995)?

Tel que présenté à l'intérieur de la problématique de cette recherche, Guba (1972) et Scriven (1995) ont identifié un malaise touchant la logique soutenant la construction du jugement. Analyser si les résultats de notre recherche appuient cette intuition est important, car cette intuition touche la vision à long terme du projet, soit mieux comprendre le malaise qu'éprouve ce domaine pour y trouver, éventuellement, une solution.

Les résultats touchant les caractéristiques des jugements et de leurs fondements (*voir* figure 4.6), tracent un portrait peu reluisant des jugements et de leurs fondements. En effet, il n'y a qu'une infime portion des rapports (2 cas) qui produisent un jugement nuancé concernant l'aspect du programme à l'évaluation tout en répondant aux questions et/ou objectifs d'évaluation. Ce constat s'ajoute au fait qu'aucun rapport n'explique la manière dont les déclarations ont été synthétisées pour construire le jugement et que plusieurs rapports émettent des recommandations qui dépassent la portée du type d'évaluation effectué (13 rapports sur les 18 ayant émis des recommandations). Enfin, peu de rapports rendent explicite la justification des choix de l'évaluation relatifs aux critères et standards utilisés. Dans la mesure où les rapports étudiés peuvent être considérés comme le reflet des pratiques évaluatives, ces constats apparaissent assez inquiétants.

Les résultats dévoilent également qu'il n'y a que la moitié des rapports étudiés qui ont été en mesure de dépasser les évidences pour produire un jugement concernant l'aspect du programme à l'évaluation. Ce constat met en doute la constitution même des rapports d'évaluation actuellement présentés. Les évaluateurs produisent-ils vraiment des évaluations ou se contentent-ils d'effectuer une enquête?

Puisque ces constats précisent un malaise au niveau du jugement, la présente recherche peut, dans une certaine mesure, confirmer les réflexions de Guba (1972) et Scriven (1995) par une méthodologie scientifique en plus de faire ressortir des pistes d'approfondissement lui étant reliées (*voir* section 5.1). Enfin, en documentant le problème de la construction des jugements que plusieurs auteurs ont abordé sans jamais avoir toutefois été analysé, elle dépasse ces intuitions pour établir les premières balises établies scientifiquement qui permettront, éventuellement, l'approfondissement de ce malaise.

Il est cependant nécessaire de nuancer ces propos en rappelant au lecteur que l'angle de traitement de cette composante réside dans la façon dont le jugement et ses fondements se présentent et non dans leur qualité. Même si Arens (2006) et Toulmin (1993) abordent bel et bien leur qualité à l'intérieur de leurs travaux, leurs contributions ont plutôt inspiré les

indicateurs témoignant de leurs caractéristiques. Il nous manquerait des appuis théoriques sur ce que signifie un jugement de qualité pour prétendre la cerner.

5.3 Quel est l'impact des résultats concernant la contribution de la modélisation à la formulation d'un jugement fondé (objectif 3) dans le domaine de l'évaluation de programme?

En ce qui a trait au troisième objectif, soit *étudier la contribution de l'application de la modélisation à la formulation d'un jugement fondé*, les résultats des analyses quantitatives et qualitatives ne révèlent pas de relation en tant que telle. Cependant, l'examen des résultats qualitatifs des jugements pour y déceler si des éléments de la modélisation y sont repris, révèle que toutes les composantes de la modélisation, hormis l'*Opérationnalisation de l'évaluation*, peuvent être réinvesties à l'intérieur d'une argumentation pour convaincre ou dissuader du bien-fondé du jugement (*voir* section 4.3.2). L'utilisation de la modélisation n'engendrerait donc pas forcément un jugement fondé mais inclurait plutôt des éléments de base pour la constitution d'une argumentation au jugement. Ces éléments, si organisés de façon à constituer une argumentation au jugement, seraient conséquemment essentiels et/ou spécifiques à l'évaluation de programme.

De ce point de vue, les prochains travaux touchant la pratique en évaluation de programme auraient tout avantage à se pencher sur ce que constitue un jugement suffisamment argumenté selon la perspective de ses clients. Leur point de vue est essentiel pour déterminer si la « force » du lien logique entre l'aspect du programme à évaluer et les critères et standards choisis (ce qui constitue l'argumentation selon Toulmin (1993) soutenu par les résultats de cette recherche (*voir* section 4.2.2) est suffisante pour permettre le passage des données au jugement. Des entrevues auprès de ceux-ci pourraient être effectuées afin d'approfondir les critères dont ils se servent pour apprécier l'argumentation et la crédibilité des jugements. Considérant que le but ultime de l'évaluation de programme est de produire un jugement suffisamment fondé pour que le client l'accepte, il est normal que le caractère appréciatif provienne de lui. Une fois ces critères cernés et étudiés, il sera plus évident de leur concéder un caractère essentiel et/ou spécifique à la pratique en évaluation de programme.

Mais qu'en est-il de l'*Opérationnalisation de l'évaluation*, la seule composante dont les éléments ne semblent pas réinvestis à l'intérieur d'une argumentation (voir section 4.3.2)? Ce constat pourrait être mis en relation avec l'analyse qualitative du contenu de cette composante (voir section 4.1.3) qui dévoile des irrégularités au niveau de son utilisation. En effet, il n'apparaît pas, tel que retrouvé à l'intérieur des rapports analysés, y avoir un consensus quant au sens exact attribué aux différents types d'évaluation et au rôle des questions/objectifs d'évaluation. Ce constat est appuyé par Scriven (1993) qui reconnaît que le rôle des objectifs n'est pas toujours clair dans le domaine de l'évaluation de programme. Ce manque de consensus expliquerait peut-être que les évaluateurs ne puissent pas la réinvestir à l'intérieur d'une argumentation pour convaincre ou dissuader du bien-fondé du jugement.

Il ne faut cependant pas oublier que les résultats à cet objectif comportent les mêmes limites que ceux présentés aux deux premiers objectifs, car ce sont à partir de ces résultats qu'ils ont été générés. Ainsi, ils se basent également sur les pratiques telles qu'écrites à l'intérieur des rapports d'évaluation et touchent uniquement les caractéristiques des jugements et non leur qualité.

CONCLUSION

L'évaluation de programme faisant face à une remise en question périodique, ce mémoire constate ce malaise et révèle plusieurs conclusions inquiétantes sur la pratique professionnelle.

Tout d'abord, les résultats obtenus au premier objectif de recherche, *établir dans quelle mesure la modélisation du processus spécifique à l'évaluation de programme se retrouve au sein des pratiques évaluatives*, mettent en lumière que la plupart des rapports ne documentent pas la *Problématique de l'évaluation*. Ils montrent également que *l'Opérationnalisation de l'évaluation*, dont l'analyse qualitative révèle des irrégularités au niveau de son utilisation, n'est pas effectuée de façon systématique

Ensuite, ceux obtenus au deuxième objectif, *caractériser les jugements qui y sont formulés*, révèlent qu'il n'y a que la moitié des rapports étudiés qui ont été en mesure de dépasser les évidences pour produire un jugement concernant l'aspect du programme à l'évaluation et qu'il y en a encore moins qui les nuancent par les limites ou forces du devis de l'évaluation. Ce constat met en doute la constitution même des rapports d'évaluation actuellement présentés. Les évaluateurs produisent-ils vraiment des évaluations ou se contentent-ils d'effectuer une enquête? Ce constat s'ajoute au fait qu'aucun rapport n'explique la manière dont les déclarations ont été synthétisées pour construire le jugement et que plusieurs rapports émettent des recommandations qui dépassent la portée du type d'évaluation effectué. Ces résultats tracent un portrait peu reluisant des jugements et de leurs fondements et précisent les intuitions de Guba (1972) et Scriven (1995) à cet égard.

Aussi, les résultats au dernier objectif de recherche, *étudier la contribution de l'application de la modélisation à la formulation d'un jugement fondé*, semble indiquer qu'il n'y a pas de relation statistiquement significative entre ces deux éléments. Cependant, une analyse plus qualitative révèle que certains éléments des composantes de la modélisation,

hormis ceux de *l'Opérationnalisation de la problématique*, peuvent être repris à l'intérieur d'une argumentation pour convaincre ou dissuader du bien-fondé du jugement.

Cette recherche apparaît ainsi novatrice de par sa méthodologie lui permettant d'approfondir le malaise touchant la connaissance et/ou application du processus spécifique à l'évaluation de programme et celle des éléments qui peuvent contribuer à générer un jugement fondé. Même si les résultats n'ont pas permis de percevoir de relation quantitative entre l'application de la modélisation et la formulation d'un jugement fondé, ils démontrent la pertinence de poursuivre les recherches sur ce que signifie un jugement de « qualité » afin de comprendre et d'améliorer les pratiques en évaluation de programme. Aussi, les constats inquiétants soulevés à l'intérieur de cette étude pourraient certainement servir de point de départ pour l'amélioration des pratiques par les praticiens ou encore pour la recherche en évaluation de programme. Enfin, rien n'empêche d'utiliser dès maintenant les critères caractérisant les jugements et leurs fondements pour encadrer et développer la pratique professionnelle de l'évaluation de programme. Les résultats peu reluisants les concernant semblent démontrer que la pratique actuelle en profiterait grandement.

Quant à la modélisation du processus spécifique à l'évaluation de programme (Hurteau et Houle, 2006), elle s'est révélé un cadre de référence de choix pour l'analyse des pratiques. Elle pourrait éventuellement être reprise par d'autres chercheurs afin de faire émerger, ultimement, l'intégralité du caractère distinctif de l'évaluation de programme.

APPENDICE A

LE DÉTAIL DES RAPPORTS D'ÉVALUATION RETENUS

Titre du rapport	Année de publication -ERIC-
2001 Literacy Program Evaluation: A report of the Evaluation of Literacy Programs in Elementary and Middle Schools.	2002
A longitudinal Evaluation of the Med-COR Program's Efforts To Improve Minority Students' Postsecondary Educational Opportunities and Health Career Participation.	2001
A study of a Reduced Class Size Program, Grades 1 & 2, 1999/2000. Evaluation Report.	2000
Addressing the Needs of Children with Autism: An Evaluation of a Gymnastics Program for Children with Autism and Related Disorders.	2000
Bilingual Education / ESL Program Evaluation, 2000-01.	2002
Children's Program Outcome Review Team: 2002 Evaluation Results.	2003
Community Jobs Outcomes Assessment & Program Evaluation.	2001
Comprehensive Evaluation of the Kentucky Extended School Services Program.	2002
Effects of a Universally Free, In-Classroom School Breakfast Program: Results from the Second Year of the Maryland Meals for Achievement Evaluation.	2000
Evaluating Two Approaches to Case Management: Implementation, Participation Patterns, Costs, and Three-Year Impacts of the Columbus Welfare-to-Work Program.	2001
Evaluation of a Ten-Year Statewide After-School Program for Struggling Learners.	2004
Evaluation of an Individualized Continuing Education Program for Physicians and Nurse Practitioners: An Example of the situational Nature of Program Evaluation.	2002

Evaluation of the “Learning by doing” Faculty Development Program for the Minnesota State Colleges and Universities.	2002
Evaluation of the Georgia CrossRoads Alternative Education Program, July 1997-June 2000.	2000
Evaluation of the Magnet Schools Assistance Program, 1998 Grantees. Final Report.	2003
Evaluation of the Public Charter Schools Program: Final Report.	2004
Final Evaluation Report: Family to Family Program	2000
High Time for High School Reform: Early findings from the evaluation of the National School District and Network Grants Program.	2003
Intervention/Prevention Program Evaluation, 1998-99.	2000
Lessons Learned from the Evaluation of Alliance+: An Internet-in-Education Professional Development Program.	2000
Making Connections: The Evaluation of the Victorian Full Service Schools Program.	2001
Native Hawaiian Safe and Drug-Free Schools and Communities Program. End of Year Evaluation Report, 1999-2000.	2000
Northeast Ohio Principals Academy Pilot Entry Year Program (Year II) Evaluation Report.	2001
Northeast Ohio Principals Academy Pilot Entry Year Program Evaluation Report.	2000
Oklahoma City Film Even Start Family Literacy Program Evaluation, 2000-2001.	2001
Program Evaluation Report for Sulphur Springs Independent School District Project STYLE.	2005
The AVID Program in AISD: Program Evaluation Report, 2000-2001.	2002
The Hebrew Academy for Special Children’s Summer Program: An Evaluation Report.	2002
The Home Visits Program at an Elementary School in California: A Summative Evaluation.	2001
The Los Angeles Jobs-First GAIN Evaluation: Final Report on a Work First Program in a Major Urban Center.	2000
The MCH Training Program: An Evaluation.	2001
The Rosen Scholars Program: A New Design for Mentoring Disadvantage Youth for Postsecondary Success.	2001
The Safe Schools/Healthy Students Initiative: Methodologies and Results in Program-Based Evaluation.	2002
Third National Even Start Evaluation: Program Impacts and Implications for Improvement.	2003
Third Party Evaluation of the effectiveness of the Structure of Intellect Model Schools Pilot Program. Year 3 Evaluation Report.	2000
Transfer: Making It Happen Program – An Evaluation	2004

Vouchers for Special Education Students: An Evaluation of Florida's McKay Scholarship Program. Civic Report.	2003
Welfare Reform: Job Access Program Improves Local Service Coordination, but Evaluation Should Be Completed.	2002
When schools Stay Open Late: The National Evaluation of the 21 st -Century Community Learning Centers Program. First Year Findings.	2003
Youth Reintegration Training and Education for Peace (YRTEP) Program: Sierra Leone, 2000-2001. Impact Evaluation.	2001

APPENDICE B

GRILLES D'ANALYSE

Grille d'analyse A Analyse du contenu lié à la <i>Description du programme</i>	p.78
Grille d'analyse B Analyse du contenu lié à la <i>Problématique de l'évaluation</i>	p.78
Grille d'analyse C Analyse du contenu lié à l' <i>Opérationnalisation de l'évaluation</i>	p.79
Grille d'analyse D Analyse du contenu lié aux <i>Choix stratégiques de l'évaluation</i>	p.80
Grille d'analyse E Analyse du contenu lié au <i>jugement et ses fondements</i> (Partie I)	p.81
Grille d'analyse E (suite) Analyse du contenu lié au <i>jugement et ses fondements</i> (Partie II)	p.82

-Grille d'analyse A-
Analyse du contenu lié à la *Description du programme*

<i>Est-ce que le rapport décrit le programme à évaluer?</i>	Oui, explicitement	Non, il faut le déduire ou ce n'est pas mentionné
1) La clientèle est mentionnée		
2) Les objectifs du programme sont mentionnés		
3) Les caractéristiques de l'intervention sont mentionnées		
4) Le cadre théorique sur lequel s'appuient les activités du programme est mentionné		

-Grille d'analyse B-
Analyse du contenu lié à la *Problématique de l'évaluation*

<i>Est-ce que le rapport présente ce à quoi l'évaluation de programme devra répondre ?</i>	Oui, explicitement ou implicitement (il est possible de le déduire de façon évidente)	Non, ce n'est pas précisé
1) L'élément déclencheur de l'évaluation est précisé		
2) Le rapport précise le type de jugement attendu par le client		
3) Le rapport précise le type d'argumentation attendu par le client		

-Grille d'analyse C-
Analyse du contenu lié à l'Opérationnalisation de l'évaluation

<i>Est-ce que le rapport opérationnalise les éléments de la Problématique de l'évaluation?</i>			
<i>Est-ce que les indicateurs de l'Opérationnalisation de l'évaluation sont présents?</i>	Oui, Explicitement	Non, il faut le déduire ou c'est absent	
1) Le/les type(s) d'évaluation effectué(s) est/sont spécifié(s)			
2) Le/les question(s) et/ou l'/les objectif(s) auquel /auxquels devra répondre l'évaluation est/sont spécifié(s)			
<i>Est-ce que ces indicateurs sont liés selon la typologie de Rossi, Lipsey, Freeman (2004)?</i>	Oui	Non	
		La typologie n'est pas respectée	Manque d'information (voir critère 1 et 2 de cette grille)
3) Le/les type(s) d'évaluation et les questions ou objectifs sont liés selon la typologie de Rossi, Lipsey, Freeman (2004)			
<i>Est-ce que ces indicateurs tiennent compte des éléments de la problématique?</i>	Oui	Non	
		Ils n'en tiennent pas tous compte	Manque d'information (voir critère 2 de la grille B ou critères 1 et 2 de cette grille))
4) Le/les type(s) d'évaluation et le/les question(s) ou objectif(s) tiennent compte du type de jugement attendu par le client			

-Grille d'analyse D-
Analyse du contenu lié aux *Choix stratégiques de l'évaluation*

Est-ce que le rapport présente la tactique utilisée pour porter un jugement?				
Est-ce que les indicateurs des choix stratégiques sont présents?	Oui, explicitement ou implicitement (il est possible de le déduire de façon évidente)		Non	
	Entièrement	En partie	Non, avec justification	Non, sans justification
1) Les critères (éléments de l'objet qui seront à l'évaluation) sont présents				
2) Les standards (points de coupure entre un jugement favorable et un jugement défavorable) sont présents				
3) Les critères ont une pondération relative qui leur est attribuée				
Est-ce que ces indicateurs tiennent compte des éléments de la problématique?	Oui		Non	
			Ils n'en tiennent pas compte	Manque d'information (voir critère 3 de la grille B ou critères 1 et 2 de cette grille)
4) Les critères et les standards tiennent compte du type d'argumentation attendu par le client				
5) Est-ce que les critères présents sont annoncés au point de départ ?				
Les critères sont présentés de façon à ce que tous les aspects de l'objet qui seront à l'évaluation soient présentés avant l'analyse des résultats.				
Certains critères émergent en cours de route, les aspects de l'objet qui seront à l'évaluation sont alors modifiés avec l'arrivée de nouvelles informations.				
Tous les critères émergent en cours de route, les aspects de l'objet qui seront à l'évaluation sont alors inclus avec l'arrivée de nouvelles informations.				
Manque d'information - voir critère 1 de cette grille				
6) Est-ce que les standards présents sont annoncés au point de départ ?				
Les standards sont présentés de façon à ce que tous les points de coupure qui détermineront le niveau de performance à atteindre soient présentés implicitement ou explicitement avant l'analyse des résultats.				
Certains standards émergent en cours de route, ces points de coupures sont alors précisés avec l'arrivée de nouvelles informations.				
Tous les standards émergent en cours de route, les points de coupures sont alors précisés avec l'arrivée de nouvelles informations.				
Manque d'information - voir critère 2 de cette grille				

-Grille d'analyse E-
Analyse du contenu lié au *jugement et ses fondements* (Partie I)

<i>D'après l'apport d'Arens (2006), comment se présente le jugement et ses fondements?</i>			
<i>Les questions trouvent-elles une réponse?</i>	Oui	Non	Manque d'information
1) Les questions (ou objectifs si absence de questions) trouvent-ils leurs réponses?			
<i>Est-ce qu'un jugement est émis et comment?</i>	Oui	Non	Manque d'information
2) Les déclarations sont-elles synthétisées pour produire un jugement global?			
3) La façon dont cette synthèse est effectuée est-elle documentée?			
<i>Les recommandations sont-elles appuyées?</i>	Oui	Non	Manque d'information
4) Des recommandations concernant la programmation sont-elles émises?			
5) Les recommandations concernent-elles l'aspect du programme à l'évaluation?			

-Grille d'analyse E-
Analyse du contenu lié au *jugement et ses fondements (Partie II)*

<i>D'après l'apport de Toulmin (1993), comment se présente le jugement et ses fondements?</i>				
<i>Le jugement est-il nuancé?</i>	Oui		Non	
6) Le jugement (ou déclarations si absence de jugement) est nuancé				
<i>Les choix stratégiques sont-ils justifiés?</i>	Oui, implicite ou explicite		Non, ces éléments ne sont pas justifiés	
	Implicite	Explicite	Ils ne sont pas justifiés	Manque d'information
7) Le rapport présente-t-il une justification concernant le choix des critères?				
8) Le rapport présente-t-il une justification concernant le choix des standards?				
<i>Sur quoi s'appuie cette justification?</i>	Éléments <i>(Il est possible d'avoir plusieurs sources de justification)</i>			
9) La justification des critères s'appuie sur...				
10) La justification des standards s'appuie sur ...				
<i>Impressions intuitives sur ce qui semble faire en sorte que le lecteur adhère au jugement ou non</i>				
11) Instinctivement, qu'est-ce qui semble faire en sorte que le lecteur adhère au jugement ou non?				

RÉFÉRENCES

- Alkin, M.C. 2003. «Evaluation Theory and Practice: Insights and New Directions». *New Direction for Evaluation*, no 97, p. 81-90.
- Alkin, M.C. 2004. *Evaluation Roots. Tracing Theorists' Views and Influences*. Thousand Oaks (Calif): Sage Publications, 424 p.
- Arens, S.A. 2005. *Understanding Practice Through Exemplary evaluation Cases: Actes du Congrès de la Société Canadienne d'Évaluation et de l'Association Américaine d'Évaluation*. (Toronto, 24-30 octobre 2005).
- Arens, S.A. 2006. L'étude du raisonnement dans les pratiques évaluatives. *Revue mesure et évaluation en éducation*, vol.29, no3, p.45-56.
- Blackburn, P.1994. *La logique de l'argumentation*. Saint-Laurent (Qué.): Éditions du renouveau pédagogique, 494 p.
- Bledsoe, K.L. et collaborateurs. 2005. *Program Theory and Theory Driven Evaluation TIG Business Meeting and Debate: Does Program Theory Really Improve Evaluation Practice? Criticisms and Questions for the Advocates: Actes du Congrès de la Société Canadienne d'Évaluation (SCE) et de l'Association Américaine d'Évaluation(AEA)*. (Toronto, 24-30 octobre 2005).
- Bouhat, C. 1993. *Guide de présentation des mémoires et thèses*. Montréal (Qué.): Université du Québec à Montréal, 110 p.
- Chen, H-T. 2005. *Practical Evaluation Program*. Thousand Oaks (Calif): Sage Publications, 292 p.
- Christie, C.A. 2003a. « What Guides Evaluation? A study of How Evaluation Practice Maps onto Evaluation Theory». *New Direction for Evaluation*, no 97, p. 7-36.
- Christie, C.A. 2003b. «Understanding Evaluation Theory and its Role in Guiding Practice: Formal, Folk, and Otherwise». *New Direction for Evaluation*, no 97, p. 91-93.
- Datta, L.E. 2003. «Important Questions, Intriguing Method, Incomplete Answers». *New Directions for Evaluation*, no 97, p. 37-46.
- Davidson, E.J. 2005. *Evaluation methodology basics: the nuts and bolts of sound evaluation*. Calif :Sage Publications, 263 p.

- De Wever, B. 2006. «Content Analysis Schemes to Analyze Transcripts of Online Asynchronous Discussion Groups: A Review ». *Computers and Education*, vol. 46, no 1, p. 6-28.
- Dubois, N. et Marceau, R. 2005. «Un état des lieux théoriques de l'évaluation : une discipline à la remorque d'une révolution qui n'en finit pas». *The Canadian Journal of Program Evaluation*, vol. 20, no 1, p. 1-36.
- Duval, J. 2005. «Étude descriptive des pratiques évaluatives : recours à la logique de l'évaluation». Mémoire de maîtrise, Montréal, Université du Québec à Montréal, 79p.
- Fetterman, D. M. 2001. *Foundations of Empowerment Evaluation*, Thousand Oaks (Calif.): Sage Publications, 173 p.
- Fink, A. 2004. *Evaluation Fundamentals Insights into the Outcomes, Effectiveness, and Quality of Health Programs*. Thousand Oaks (Calif.): Sage Publications, 280 p.
- Fournier, D.M. 1995. «Establishing Evaluative Conclusions: A Distinction Between General and Working Logic». *New Direction for Evaluation*, no 68, p. 15-32.
- Fournier, D. M. et Smith, N. L. 1993. «Clarifying the Merits of Argument in Evaluation Practice». *Evaluation and Program Planning*, vol.16, no 4, p.315-323.
- Gaudreau, L. 2001. *Évaluer pour évoluer : les indicateurs et les critères*. Montréal (Qué.): Les éditions logiques, 71 p.
- Guba, E.G. 1972. «The Failure of Educational Evaluation ». In *Evaluating Action Programs : Readings in Social Action and Education*, sous la dir. de Weiss C. H., p. 251-266. Boston: Allyn et Bacon.
- Guba, E.G. et Lincoln, Y.S. 1989. *Fourth Generation Evaluation*. Newbury Park (Calif.): Sage Publications, 294 p.
- House, E.R. 1980. *Evaluating With Validity*. Thousand Oaks (Calif.): Sage Publications, 294 p.
- 1995. «Putting things together coherently : Logic and justice». *New directions for evaluation*, no.68, p.33-48
- House, E.R. et Howe, K.R. 1999. *Values in evaluation and social research*. Thousand Oaks (Calif.): Sage Publications, 178p.
- Hurteau, M. 1991. «Strategic Choices in Program Evaluation». In *Evaluation Methods Sourcebook*, sous la dir. De Love, A. J. Ottawa (Ont.): Les publications de la société canadienne d'évaluation, p.106-125.

- Hurteau M. et Houle S. 2005a. «Évaluation de programme et recherche évaluative: des activités distinctes» *Mesure et Évaluation en Éducation*, vol. 28, no 3, p. 83-95.
- 2005b. *Identifying a Core Body of Knowledge for Evaluators. Actes du congrès annuel de l'Association Américaine d'Évaluation (AEA)*. (Toronto, 27 Octobre 2005) Acte du congrès électronique: <http://lg.evaluationcanada.ca>.
- Hurteau, M., Houle, S. 2006. *Distinction entre l'acte spécifique d'évaluer et ses modalités de soutien*. Congrès de l'ADMEE Europe, Luxembourg.
- Hurteau, M., Lachapelle, G. et Houle, S. 2006. «Comprendre les pratiques évaluatives afin de les améliorer : la modélisation du processus spécifique à l'évaluation de programme ». *Mesure et Évaluation en Éducation*, vol.29, no 3, p. 27-44.
- Hurteau, M. et Nadeau, M.A. 1987. «Évaluation de programmes, recherche évaluative et recherche.». *Revue canadienne de psycho-éducation* , vol. 12, no 3, p. 404-416.
- Kaplan S.A. et Garrett, K.E. 2005. «The Use of Logic Models by Community-Based Initiatives». *Evaluation and Program Planning*, vol. 28, p.167-172.
- Karsenti, T. et Savoie-Zajc, L. 2000. *Introduction à la recherche en éducation*. Sherbrooke (Qué.): Editions du CRP, 350 p.
- King, J.A. 2003. «The Challenge of Studying Evaluation Theory». *New Direction for Evaluation*, no 97, p. 57-67.
- Lipsey, M.W. 1988. «Practice and Malpractice in Evaluation Research». *Evaluative Practice*, vol. 9, no 4, p. 5-24.
- Madaus, G.F. et Stufflebeam, D.L. 2004. « Program Evaluation: A Historical Overview ». In *Evaluation Models : Viewpoints on Educational and Human Services Evaluation* (2e éd.), sous la dir. de Stufflebeam, D.L. Madaus, G.F. et Kellagan T., p. 3-19. Boston: Kluwer Academic Publishers.
- Mathison, S. et collaborateurs. 2004. *Encyclopedia of Evaluation*. Thousand Oaks (Calif.): Sage Publications, 481 p.
- McCarthy, T.A. 1973. «A theory of communicative competence». *Philosophy of the Social Sciences*, no 3, p.135-156.
- McDavid, J.C. 2001. «Program Evaluation In British Columbia in a Time of Transition: 1995-2000». *The Canadian Journal of Program Evaluation*, vol.16, p. 3-28.
- McLaughlin, J. et Jordan, G. 1999. «Logic Models: A Tool for Telling your Program's Performance Story ». *Evaluation and Program Planning*, vol. 22, p. 65-72.

- Millar, A., Simeone, R. et Carnevale, J. 2001. «Logic Models: A Systems Tool for Performance Management ». *Evaluation and Program Planning*, vol.24, 73-81.
- Neuendorf, K.A. 2002. *The Content Analysis Guidebook*. Thousand Oaks (Calif.): Sage Publications, 308 p.
- Patton, M.Q. 2001. «Evaluation, Knowledge Management, Best Practices, and High Quality Lessons Learned ». *American Journal of Evaluation*, vol. 22, no 3, p. 329-336.
- Redding, P. 1989 « “Habermas”, theory of Argumentation ». *Journal of Value Inquiry*, no 23, p.15-32.
- Rossi, P. H., Lipsey, M.W. et Freeman, H.E. 2004. *Evaluation : A Systematic Approach*, Thousand Oaks (Calif.): Sage Publications, 470 p.
- Rourke, L. et Anderson, T. 2004. «Validity in Quantitative Content Analysis». *Educational Technology Research and Development*, vol. 52, no 1, p. 5-17.
- Sanders, J. R. et collaborateurs (Joint Committee on Standards for Educational Evaluation, American Association of School Administrators). 1994. *The Program Evaluation Standards: How to Assess Evaluations of Educational Program*. Thousand Oaks (Calif.): Sage publication, 222 p.
- Sanders, J.R. 2001. « A Vision for Evaluation ». *American Journal of Evaluation*, vol. 22, no 3, p. 263-266.
- Scriven, M.1980. *The Logic of Evaluation*. Inverness (Calif.): Edgepress.125 p.
- 1991. *Evaluation Thesaurus* (4^e éd.). Newbury Park (Calif.) : Sage Publications, 408 p.
- 1993. «Hard-won Lessons in Program Evaluation». *New Directions for Evaluation*, no 58, p.1-103.
- 1995. «The Logic of Evaluation and Evaluation Practice ». *New Direction for Evaluation*, no 68, p.49-70.
- 2001. «Evaluation: Future Tense». *American Journal of Evaluation*, vol. 22, no 3, p. 301-307.
- 2006. «Value-Driven Evaluation» : Actes du congrès annuel de la Société Américaine d'Évaluation (Portland, novembre 2006). Acte du congrès électronique : <http://lg.evaluationcanada.ca>.
- Secrétariat du Conseil du Trésor du Canada. 2004. *Examen de la qualité des évaluations dans les ministères et les organismes*. Document interne, Gouvernement du Canada (http://www.tbs-sct.gc.ca/eval/pubs/rev-exam_f.asp).

- Shadish, W.E., Cook, T.D. et Leviton, L.C. 1991. *Foundations of Program Evaluation: Theories of Practice*. Thousand Oaks (Calif.): Sage Publications, 529 p.
- Shadish, W.R., Cook, T. D. et Campbell, D.T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston : Houghton Mifflin, 623 p
- Shaw, I., Greene, J.C. et Mark, M.M. 2006. *Handbook of Evaluation: Policies, Programs and Practices*. London: Sage Publications, 608 p.
- Stake, R.E. 2004. *Standards-Based and Responsive Evaluation*. Thousand Oaks (Calif.): Sage Publications, 329 p.
- Stufflebeam, D.L. 2001. «Evaluation Models ». *New directions for evaluation*, no 89, p.7-98.
- Suchman, E. 1967. *Evaluative Research : Principles and Practice in Public Service and Social Action Program*. New York: Sage Publications, 186 p.
- Taylor, P.W. 1973. *Normative Discourse*. Westport (Conn.): Greenwood Press, 360 p.
- Toulemonde, J. 2005. *Appropriation des résultats de l'évaluation : Leçons de la pratique en Région Limousin. Acte du colloque de la Société Française de l'Évaluation*. (Lille, mai 2005).
- Toulmin, S.E. 1993. *Les usages de l'argumentation*. Paris : Presses universitaires de France, 325 p.
- Toulmin S.E., Rieke, R.D. et Janik, A. 1984. *An Introduction to Reasoning*. New York: Macmillan, 435 p.
- Trelogan, T.K. 2001. *Arguments and their evaluation*. Department of philosophy. Colorado: University of Northern.
- Virtanen, P. et Uusikylä, P. 2002. «Exploring the Missing Links Between Causes and Effects». Actes de la Conférence annuelle de la Société Européenne d'Évaluation. (Séville, octobre 2002).
- Weiss, C.H. 1998. *Evaluation: Methods for Studying Programs and Policies*. New Jersey: Pearson Education Canada, 372 p.
- Wholey, J.S., Hatry H.P. et Newcomer, K.E. 2004. *Handbook of Practical Program Evaluation*. San Francisco : Jossey-Bass, 720 p.